

2-РО ИЗДАНИЕ,  
АКТУАЛИЗИРАНО И  
РАЗШИРЕНО

# DATA DRIVEN CONSTRUCTION

## НАВИГИРАНЕ В ЕРАТА НА ДАННИТЕ В СТРОИТЕЛНАТА ИНДУСТРИЯ

С ПРИМЕРИ ЗА УПОТРЕБА  
НА AI И LLM

Artem Boiko



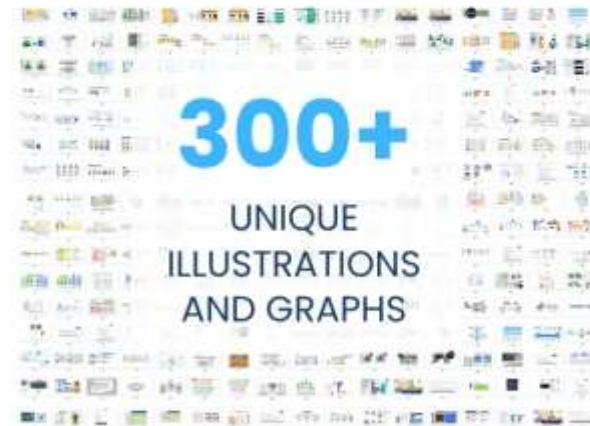
**100+**

KEY DATA-  
RELATED  
TOPICS



**50+**

PRACTICAL  
PROBLEMS  
SOLVED



**300+**

UNIQUE  
ILLUSTRATIONS  
AND GRAPHS

# DATA-DRIVEN CONSTRUCTION

Навигация в ерата на данните  
в строителната индустрия

Второ издание, преработено и допълнено

ARTEM BOIKO

**“**

"Бойко е Джеймс Карвил на ИТ - в многократно цитираната от него фраза "It's the economy, stupid" (Икономиката е глупава) само една дума трябва да се замени с тази известна книга. "Това са данните, глупако". (А за да се намери път във вселената на данните, една поговорка на древните римляни, датираща от гръцки, е валидна и днес: "Navigare necesse est". Авторът навигира своите читатели през всички дълбини и плитчини на океана от данни с уверена ръка и непоколебим компас, да не говорим за изчерпателен исторически подход и не на последно място за изключително оригинални графики и добро чувство за хумор, което се вижда не само на втори поглед. Международният отзив от книгата на Бойко варира от еуфорично одобрение до доста жълчен скептицизъм, което е допринесло за второто немско издание на книгата. Бойко е оригинален и недогматичен мислител на данни. Той представя на читателя вълнуващи прозрения и винаги смели, дори провокативни тези, които вдъхновяват за по-нататъшни размисли. Отлично лекарство за немската болест на латентния консенсуализъм. Между другото, горната латинска поговорка има допълнение: "vivere non est necesse". Тя не важи за подхода на Бойко към света на данните - данните живеят и техният живот е необходим, да не кажем решаващ."

- **Д-р Буркхард Талебитари**, редактор на свободна практика - включително за списанието: BIM, издавано ежегодно от Ernst & Sohn от 2013 г. насам.

**"Книгата на Артьом Бойко е крайъгълен камък за демократизирането на цифровизацията в строителната индустрия - и истинска промяна за малките и средните предприятия (МСП). Особено новаторска: чрез използването на съвременни инструменти с отворен код с нисък код и без код компаниите вече могат ефективно да интегрират данните в своите бизнес процеси и да ги анализират с печалба - без задълбочени познания по програмиране. Това прави излишно скъпото използване на тромави търговски софтуерни пакети. Тази книга е призив за действие! Тя е ценно ръководство за всеки, който не само иска да разбере цифровата трансформация в строителния бранш, но и иска активно да я оформи - прагматично, ефективно и с поглед към бъдещето. Сега е моментът да работим заедно, за да споделим тези знания и устойчиво да увеличим производителността на строителната индустрия."**

**-Д-р Михаел Макс Бюлер**, професор по строителен менеджмънт в HTWG Констанц, съсобственик в GemeinWerk Ventures и независим директор в DevStream.

**"Книгата DataDrivenConstruction е една от първите стъпки отвъд границите на обичайния свят на строителите, с техните сложни системи за проектиране и управление, когато, изглежда, сложността и насищността с данни дори не дава възможност за радикално опростяване и по-голяма прозрачност на работата със строителни данни. В книгата си Артем показва на прост език какви възможности откриват пред нас съвременните технологии за работа с данни и буквально дава конкретни стъпки, които можете веднага да приложите в работата си. Призовавам всички, които искат да разберат докъде ще стигнат системите за автоматизация в строителството, да се запознаят внимателно с тази книга, за да осъзнаят, че революцията на данните в строителството вече чука на вратата ни. Сега тя представлява интерес само за маниаките, но след няколко години, подобно на BIM, подобни подходи и софтуер ще бъдат повсеместни!"**

**- Ихор Рогачев**, ръководител на Центъра за компетентност IMT, BIM и цифрова трансформация в RGD и основател на InfraBIM.Pro.

**"Горещо препоръчвам книгата DataDrivenConstruction, в която, както се казва в заглавието, се разглежда подход за управление на информацията, базиран на данни, за AECO. В момента я използвам, за да помогна за започването на редица дискусии с различни групи. Смятам, че тя е много достатъчен справочник. Освен задълбочен преглед на историческия контекст на инструментите в AECO, данните и въвеждане на няколко ключови технологии, книгата съдържа редица много полезни диаграми, които очертават обхвата на източниците на данни и артефактите на крайния потребител с примерни работни процеси. Струва ми се, че това са видовете диаграми, от които се нуждаем повече при разработването и наблюдението на информационните стратегии и допринасят за ВЕР - определят цялостния модел на данните на предприятието, върху който може да се наложи границата за PIM и AIM."**

**- Пол Рансли**, главен консултант в Acumen и инженер по системна интеграция в Transport for London.

**"Ако "данните са новият петрол", трябва да се научим да ги определяме, намираме, добиваме, преработваме, за да ги направим ценни. Намерих книгата DataDrivenConstruction за много информативна и проницателна. Книгата представя полезна историческа справка и обяснява работата с данни на разбираем език. За тези, които се интересуват от цифровата трансформация, тя дава добра представа за данните - как работят, как са структурирани и как могат да бъдат използвани."**

**- Ралф Монтею**, директор в ArcDox, директор на срещата на върха на координаторите на BIM и председател на Националния огледален комитет за BIM в Националния орган по стандартизация на Ирландия.

**"Както беше подчертано в книгата, информацията е изключително важен актив за строителния сектор и наличието на достъпна форма значително улеснява вземането на точни решения и ускорява сроковете на проектите. Книгата предлага неутрален и ефективен подход за достъп и използване на този източник при вземането на решения. Методологията, представена в книгата, използва съвременен подход, който съчетава програмиране, базирано на изкуствен интелект, с достъпни инструменти с отворен код. Като използва силата на изкуствения интелект и използва софтуер с отворен код, методологията има за цел да повиши автоматизацията, да оптимизира процесите и да насычи достъпността и сътрудничеството в областта. Езикът на книгата е ясен и лесен за следване."**

**- Д-р Салих Офлуоглу**, декан на Факултета по изящни изкуства и архитектура в университета "Антalia Билим" и организатор на Евразийския форум за BIM.

**"Всичко, което мога да кажа, е WOW! Начинът, по който сте включили историята, LLM, графиките и цялостната лекота на разбиране на вашите точки, е наистина забележителен. Потокът на книгата е невероятен. В тази книга има толкова много блестящи аспекти; тя наистина променя правилата на играта. Тя е чудесен източник на информация и аз ви поздравявам за усилията и страсти, които сте вложили в нея. Поздравявам ви за създаването на такова забележително произведение. Мога да продължа, но е достатъчно да кажа, че съм невероятно впечатлен!"**

**- Наташа Принслу**, ръководител на цифровата практика в energylab\_

**"За всеки в строителния бранш, от начинаещи до опитни професионалисти, тази книга променя правилата на играта! Тя не е типичното прашно четиво - пълна е с прозрения, стратегии и нотка хумор, за да ви държи в напрежение. Тя обхваща еволюцията на използването на данни в строителството - от древните методи за записване на данни до най-съвременните цифрови технологии. Сякаш се возите на машина на времето през еволюцията на данните в строителството. Независимо дали сте архитект, инженер, ръководител на проект или анализатор на данни, това изчерпателно ръководство ще промени коренно начина, по който подхождате към проектите. Пригответе се да оптимизирате процесите, да подобрите вземането на решения и да управлявате проекти както никога досега!"**

**- Пиерпаоло Вергати**, преподавател в университета "Сапиенца" в Рим и старши мениджър на строителни проекти във Fintecna.

"Прочетох книгата на един дъх, за по-малко от 6 часа. Качеството на изработка на книгата е отлично, пътна гланцирана хартия, цветови гами, приятен шрифт. Големият брой практически примери за работа с LLM, специфични за строителния бранш, ще ви спестят месеци, ако не и години, самообучение. Примерите за работа са много разнообразни, варират от прости до сложни, без да се налага да закупувате сложен и скъп софтуер. Книгата ще позволи на собствениците на всяко предприятие в строителния бранш да погледнат по нов начин на своята бизнес стратегия, цифровизация и перспективи за развитие. А за по-малките компании - да повишат ефективността си с достъпни и безплатни инструменти."

- **Михаил Косарев**, лектор и консултант по цифрова трансформация в строителната индустрия в TIM-ASG.

Книгата "DATA DRIVEN CONSTRUCTION" променя правилата на играта за всеки, който се интересува от това накъде се движи строителната индустрия в ерата на данните. Артем не просто надраска повърхността, а навлиза в дълбочина на текущите развития, предизвикателства и обещаващи възможности в строителството. Това, което отличава тази книга, е нейната достъпност - Артем обяснява сложни идеи с помощта на относителни аналогии, които правят съдържанието лесно за възприемане. Смятам, че книгата е изключително информативна и същевременно увлекателна. В обобщение, Артем е създал ценен ресурс, който не само информира, но и вдъхновява. Независимо дали сте опитен професионалист, или новак в строителството, тази книга ще разшири гледната ви точка и ще задълбочи разбирането ви за това накъде върви индустрията. Силно препоръчителна!"

- **Моаяд Салех**, архитект и мениджър внедряване на BIM в TMM GROUP Gesamtplanungs GmbH.

"Бих искал да кажа, че "Data-Driven Construction" заслужава да бъде преподавана като учебник в университетите и е книга, която ще има ценен принос към развиващата се област на BIM. Книгата "Data-Driven Construction" съдържа технически речник, който много добре обяснява понятията. Теми, които са изключително трудни за обяснение, са направени прости и разбираеми с много красив визуален език. Смятам, че това, което се цели да бъде обяснено във визуалните материали, трябва да бъде изразено пред читателя, макар и накратко. Разбираемостта на някои визуализации, с други думи, четенето на визуализацията изисква отделна информация. Бих искал също така да кажа, че с удоволствие представям ценните произведения на Артьом Бойко в своите лекции и семинари в университетите."

- **Д-р Едиз Язчиоглу**, собственик на ArchCube и преподавател по управление на строителни проекти в катедрата по архитектура в Истанбулския технически университет и в университета Medipol.

"Data Driven Construction" нагледно представя основите на информационно базираната работа със строителни данни. Книга, която се занимава с информационните потоци и фундаменталните икономически концепции и по този начин се отличава от другите книги за BIM, защото не само представя гледната точка на производителя на софтуер, но и се опитва да предаде фундаментални концепции. Книга, която си заслужава да бъде прочетена и видяна."

- **Якоб Хирн**, главен изпълнителен директор и съосновател на Build Informed GmbH, и инициатор на Форума за иновации "На върха с BIM".

"Данните са новият петрол", както се казва, така че техните търсачи или миньори трябва да разполагат с правилните инструменти и начин на мислене, за да извлекат стойност от този ресурс на 21-ви век. Строителният бранш твърде дълго време се намираше на хълзгавия склон на процесите, базирани на "3D информация", при които изпълнението на проекта се основава на изпечена информация от някой друг (например вече са начертали кръговата или стълбовидната диаграма), докато основните "данни" (например сировата електронна таблица) са способни да осигурят много повече, особено защото сливането на множество данни и изкуственият интелект носят неограничен потенциал. Ако доставяте (или преподавате/изследвате) строителство, тази книга е най-добрят ви - и засега единствен - ресурс за ориентиране в света, управляван от данни, в който сме попаднали."

- **Д-р Зулфикар Адаму**, доцент по стратегически информационни технологии в строителството в LSBU, Великобритания.

"Data-Driven Construction" на Артем Бойко е впечатляващ труд, който предлага солидна основа за строителния бранш във времена на постоянно развиващи се технологии и информационни възможности. Бойко успява да представи сложните теми по разбирамен начин, като същевременно въвежда визионерски идеи. Книгата е добре обмислен сборник, който не само акцентира върху настоящите разработки, но и дава перспектива за бъдещите иновации. Тя се препоръчва на всички, които искат да се запознаят с планирането и изпълнението на строителството, основано на данни."

- **Маркус Айбергер**, преподавател в Университета за приложни науки в Щутгарт, старши мениджър на проекти и заместник-ръководител на клон в Konstruktionsgruppe Bauen, член на управителния съвет на сдружението BIM Cluster Baden-Württemberg.



Второ издание, март 2025 г.  
© 2025 | Артем Бойко | Карлсруе

ISBN 978-3-912002-05-8



Артем Бойко Авторско право

[boikoartem@gmail.com](mailto:boikoartem@gmail.com)  
[info@datadrivenconstruction.io](mailto:info@datadrivenconstruction.io)

Никоя част от тази книга не може да бъде възпроизвеждана или предавана под каквато и да е форма или по какъвто и да е начин, електронен или механичен, включително фотокопиране, запис или каквато и да е система за съхранение и извлечане на информация, без писменото разрешение на автора - освен за нетърговско разпространение в непроменен вид. Книгата се разпространява безплатно и може да бъде свободно споделяна с други потребители за лични, образователни или изследователски цели, при условие че се запазят авторството и препратките към оригинала. Авторът запазва всички неимуществени права върху текста и не дава никакви изрични или подразбиращи се гаранции. Споменатите в книгата фирми, продукти и имена може да са фiktивни или използвани като примери. Авторът не носи отговорност за последиците от използването на дадената информация. Информацията, съдържаща се в книгата, се предоставя "такава, каквато е", без гаранция за пълнота или уместност. Авторът не носи отговорност за случайни или последващи щети, произтичащи от използването на информацията, кода или програмите, съдържащи се в тази книга. Представените в тази книга примерни кодове са предназначени само за образователни цели. Читателите ги използват на свой риск. Авторът препоръчва всички софтуерни решения да бъдат тествани преди използването им в производствена среда. Всички търговски марки и имена на продукти, споменати в текста, са търговски марки, регистрирани търговски марки или марки за услуги на съответните компании и са собственост на съответните им собственици. Използването на тези имена в книгата не означава каквато и да е връзка с техните собственици или одобрение от тяхна страна. Споменаването на продукти или услуги на трети страни не представлява препоръка и не предполага одобрение. Имената на компании и продукти, използвани в примерите, може да са търговски марки на техните собственици. Връзките към уебсайтове на трети страни са предоставени за удобство и не означават, че авторът одобрява информацията, предоставена на тези сайтове. Всички цитирани статистически данни, цитати и проучвания са актуални към момента на написване на статията. Данните могат да се променят с течение на времето.

Тази книга се разпространява под лиценза Creative Commons Признание-Некомерсиално-Недеривативи 4.0 International (CC BY-NC-ND 4.0). Можете да я копирате и разпространявате за нетърговски цели, при условие че запазите авторството и не правите промени.



CC BY-NC-ND 4.0

© 2024 Артьом Бойко. Първо издание.

© 2025 Артьом Бойко. Второ издание, преработено и допълнено.

Всички права запазени.

## ПРЕДГОВОР КЪМ ВТОРОТО ИЗДАНИЕ

Тази книга е резултат от оживен диалог с професионалната общност. Тя се основава на многобройни професионални дискусии за управлението на данни в строителния бранш, които се провеждаха на различни професионални платформи и платформи на социалните медии. Тези дискусии станаха основа за статии, публикации и визуални материали, които предизвикаха широк отзив в професионалната общност. Съдържанието на автора привлича милиони гледания всяка година в различни платформи и на различни езици, като обединява професионалистите в областта на цифровизацията на строителството.

В рамките на една година след публикуването на първото издание книгата е поръчана от експерти от повече от 50 държави - от Бразилия и Перу до Мавриций и Япония. Второто издание на книгата, което сега държите в ръцете си, е преработено и разширено въз основа на отзивите на експертите, критиките към първото издание и дискусиите в професионалните среди. Благодарение на отзивите второто издание е значително разширено: добавени са нови глави за CAD (BIM) технологиите и създаването на ефективни ETL-процеси. Бroat на практическите примери и казуси също е значително увеличен. Особено ценна е обратната връзка от ръководители на строителния бранш, консултантски компании и големи ИТ-компании, които се обърнаха към автора с въпроси за цифровизацията и оперативната съвместимост както преди, така и след публикуването на първата версия на книгата. Много от тях вече са приложили подходите, описани в книгата, или планират да го направят в близко бъдеще.

Вие държите в ръцете си книга, създадена чрез дискусии и активен обмен на мнения. Напредъкът се ражда в диалога, в сблъсъка на мнения и в отвореността към нови подходи. Благодарим ви, че сте част от този диалог. Вашата конструктивна критика е в основата на бъдещи подобрения. Ако открите грешки в текста или искате да споделите идеи и предложения, всяка обратна връзка е добре дошла. Данните за контакт са посочени в края на книгата.

## ЗАЩО КНИГАТА Е БЕЗПЛАТНА?

Тази книга е замислена като отворен образователен ресурс, целящ да разпространи съвременни подходи за управление на данни в строителната индустрия. Първата версия на книгата послужи като основа за събиране на коментари и предложения от професионалната общност, които позволиха да се подобрят структурата и съдържанието на материала. Всички коментари, предложения и идеи бяха внимателно анализирани и включени в настоящата преработена версия. Целта на книгата е да помогне на строителните специалисти да разберат важността на работата с данни: систематично, съзнателно и с оглед на дългосрочната стойност на информацията. Авторът е събрали примери, илюстрации и практически наблюдения от повече от 10 години работа в областта на цифровизацията на строителството. По-голямата част от този материал е породена от реални проекти, дискусии с инженери и строители, участие в международни инициативи и обучителни семинари. Книгата е опит да се структурира натрупаният опит и да се сподели по достъпен начин. Ако искате да подкрепите по-нататъшното разпространение на идеите на книгата и да получите удобен формат за четене, работа с примери и визуални материали - можете да закупите [печатна версия](#).

## ПРАВА НА ИЗПОЛЗВАНЕ

Всички материали, илюстрации и фрагменти от тази книга могат да бъдат възпроизвеждани, цитирани или използвани в какъвто и да е формат и на каквато и да е медия, при условие че източникът е посочен като Artem Boiko, а заглавието на книгата е "Data-Driven Construction". Благодарим ви за уважението към труда и разпространението на знания.

С искрена благодарност посвещавам тази книга на семейството си, което от ранна възраст ми вдъхна дълбока любов към строителството, на родния ми миньорски град за уроците по устойчивост и на съпругата ми геодезистка, чиято неизменна подкрепа е моето постоянно вдъхновение.

## **ЗА КОГО Е ПРЕДНАЗНАЧЕНА ТАЗИ КНИГА**

Написана на достъпен език, тази книга е предназначена за широк кръг читатели в строителния бранш - от студенти и начинаещи, които искат да разберат основите на съвременните строителни процеси, до професионалисти, които се нуждаят от съвременна методология за управление на данните в строителството. Независимо дали сте архитект, инженер, бригадир, строителен мениджър или анализатор на данни, това изчерпателно ръководство с много уникални илюстрации и графики предлага ценни идеи за това как да използвате данните в бизнеса, за да оптимизирате и автоматизирате процесите, да подобрите вземането на решения и да управлявате строителни проекти на различни нива с помощта на съвременни инструменти.

Книгата е изчерпателно ръководство, което съчетава теоретични основи и практически препоръки за интегриране на техники за управление на данни в строителните процеси. Книгата се фокусира върху стратегическото използване на информацията за оптимизиране на дейността, автоматизиране на процесите, подобряване на процеса на вземане на решения и ефективно управление на проекти с помощта на съвременни цифрови инструменти.

Тази книга обхваща теоретичните и практическите аспекти на работата с информация в строителната индустрия. Чрез подробни примери тя разглежда методологията за параметризиране на задачите, събиране на изисквания, обработка на неструктурирани и многоформатни данни и превръщането им в ефективни решения за строителните компании.

Читателят последователно преминава през пътя от формирането на изисквания и разработването на основни модели на данни до по-сложни процеси на интегриране на разнородни източници на информация, създаване на ETL-процеси, изграждане на информационни тръбопроводи и модели за машинно обучение. Последователният подход позволява да се демонстрират нагледно механизмите на организация и автоматизация на бизнес процесите и системите за подпомагане на вземането на решения в строителния бранш. Всяка част на книгата завършва с практическа глава, съдържаща инструкции стъпка по стъпка, които позволяват незабавно прилагане на придобитите знания в реални проекти.

## РЕЗЮМЕ НА ЧАСТИ ОТ КНИГАТА

Тази книга е структурирана около концепцията за трансформация на данните във веригата на стойността: от събирането на данни и осигуряването на качество до аналитичната обработка и извлечането на ценни практически решения с помощта на съвременни инструменти и методологии.

**Част 1: Цифрова еволюция в строителството** - проследява историческата трансформация на управлението на данни от глинени плочки до съвременни цифрови системи, като анализира появата на модулни системи и нарастващото значение на информационната цифровизация в контекста на индустриалните революции.

**Част 2: Информационни предизвикателства пред строителната индустрия** - разглежда проблемите с фрагментацията на данните, "информационните силози", въздействието на подхода HiPPO вземането на решения и ограниченията на патентованите формати, като предлага да се обмисли преминаването към екосистеми с изкуствен интелект и LLM.

**Част 3: Систематизиране на данните в строителството** - формира типология на данните в строителството, описва методите за тяхното организиране, интегриране с корпоративните системи и обсъжда създаването на центрове за компетентност за стандартизиране на информационните процеси.

**Част 4: Осигуряване на качеството на данните** - разкрива методологии за превръщане на разнородна информация в качествени, структурирани данни, включително извлечане на данни от различни източници, валидиране и моделиране с помощта на LLM.

**Част 5: Изчисления на разходите и времето** - разглежда цифровизацията на изчисленията на разходите и планирането, автоматизацията на получаването на обеми от CAD (BIM) модели, технологиите за 4D-8D моделиране и ESG изчисленията на строителни проекти.

**Част 6: CAD и BIM** - критичен анализ на еволюцията на технологиите за проектиране, проблемите с оперативната съвместимост на системите, тенденциите към отворени формати за данни и перспективите за прилагане на изкуствен интелект в проектирането.

**Част 7: Анализ на данни и автоматизация** - разглежда принципите на визуализация на информация, ключови показатели за ефективност, ETL процеси, инструменти за оркестрация на работни потоци и прилагане на езикови модели за автоматизиране на рутинни задачи.

**Част 8: Съхраняване и управление на данни** - разглежда формати за съхранение на данни, концепции за хранилище на данни и езеро от данни, принципи за управление на данни и нови подходи, включително векторни бази данни и методологиите DataOps и VectorOps.

**Част 9: Големи данни и машинно обучение** - фокусира се върху прехода към обективен анализ, основан на исторически данни, интернет на нещата на строителните обекти и прилагането на алгоритми за машинно обучение за прогнозиране на разходите и сроковете на проектите.

**Част 10: Строителната индустрия в ерата на цифровите данни** - представя поглед към бъдещето на строителната индустрия, като анализира прехода от причинно-следствен анализ към работа с корелации, концепцията за "Uberisation" на строителството и стратегиите за цифрова трансформация.

# What is meant by **data-driven construction** ?



# ВЪВЕДЕНИЕ

Колко дълго вашата компания може да остане конкурентоспособна в свят, в който технологиите се развиват бързо и всеки аспект на бизнеса - от планирането на времето и изчисляването на разходите до анализа на риска - се автоматизира от модели за машинно обучение?

Строителният бранш, който съществува толкова дълго, колкото и самото човечество, е на прага на революционни промени, които обещават да променят изцяло начина, по който мислим за традиционното строителство. Вече и в други сектори на икономиката цифровизацията не просто променя правилата, но и безмилостно изтласква от пазара компаниите, които не са успели да се адаптират към новата среда за обработка на данни и не са в състояние да подобрят скоростта на вземане на решения (фиг. 1).



**Фиг. 1** Бързината на вземане на решения в строителния бранш зависи от човешкия фактор по-често, отколкото в други отрасли.

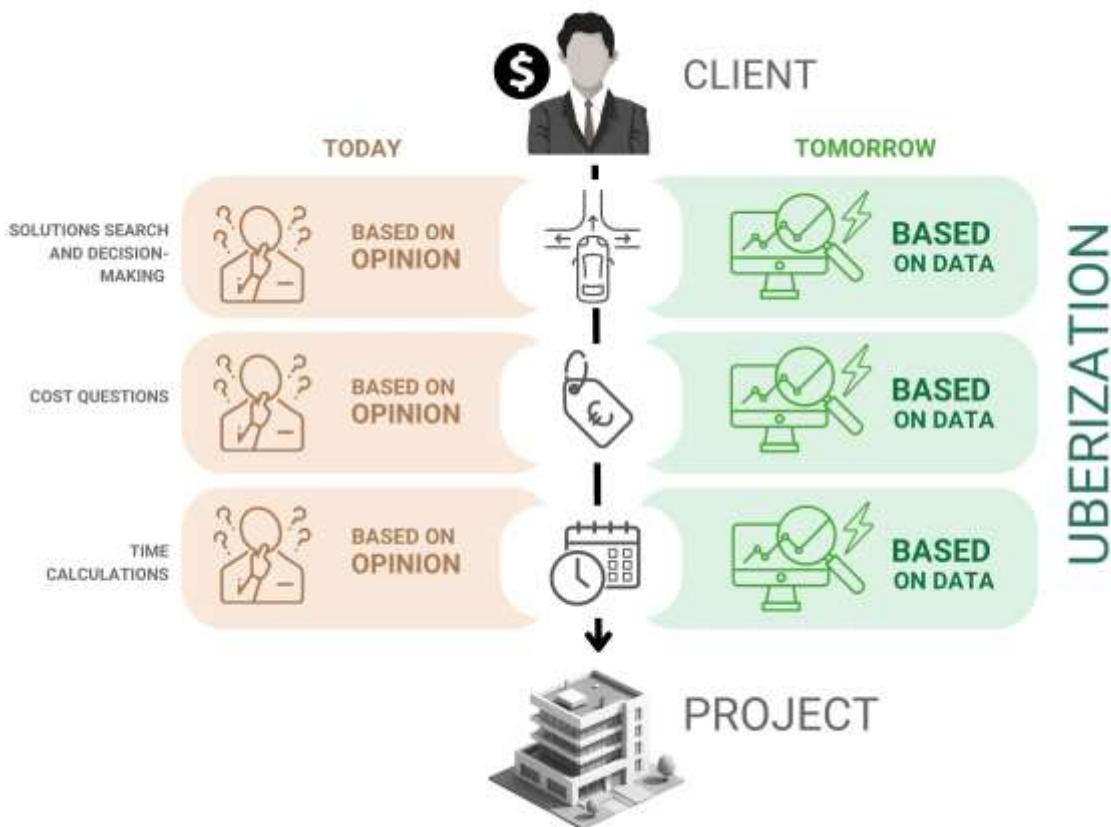
Банковото дело, търговията на дребно, логистиката и агробизнесът бързо преминават към пълна цифровизация, в която неточностите и субективните мнения вече нямат място. Съвременните алгоритми са в състояние да анализират огромни количества данни и да предоставят на клиентите точни прогнози - независимо дали става въпрос за вероятността за погасяване на кредит, оптимални маршрути за доставка или прогнозиране на риска.

Строителството е една от последните индустрии, които правят неизбежния преход от решения, базирани на мненията на високоплатени специалисти, към решения, базирани на данни. Този преход се обуславя не само от новите технологични възможности, но и от повишенните изисквания на пазара и клиентите за прозрачност, точност и бързина.

Роботизацията, автоматизацията на процесите, отворените данни и базираните на тях прогнози - всичко това вече не е просто възможност, а неизбежност. Повечето компании в строителния

бранш, които доскоро отговаряха пред клиента за изчисляването на обема, разходите, времето за изпълнение на проектите и контрола на качеството, сега рискуват да се превърнат в обикновени изпълнители на поръчки, без да вземат ключови решения (фиг. 2).

С напредването на изчислителната мощ, алгоритмите за машинно обучение и демократизирания достъп до данни вече е възможно автоматично да се комбинират данни от различни източници, което дава възможност за по-задълбочен анализ на процесите, прогнозиране на рисковете и оптимизиране на разходите на етапите на обсъждане на даден строителен проект. Тези технологии имат потенциала радикално да подобрят ефективността и да намалят разходите в целия сектор.



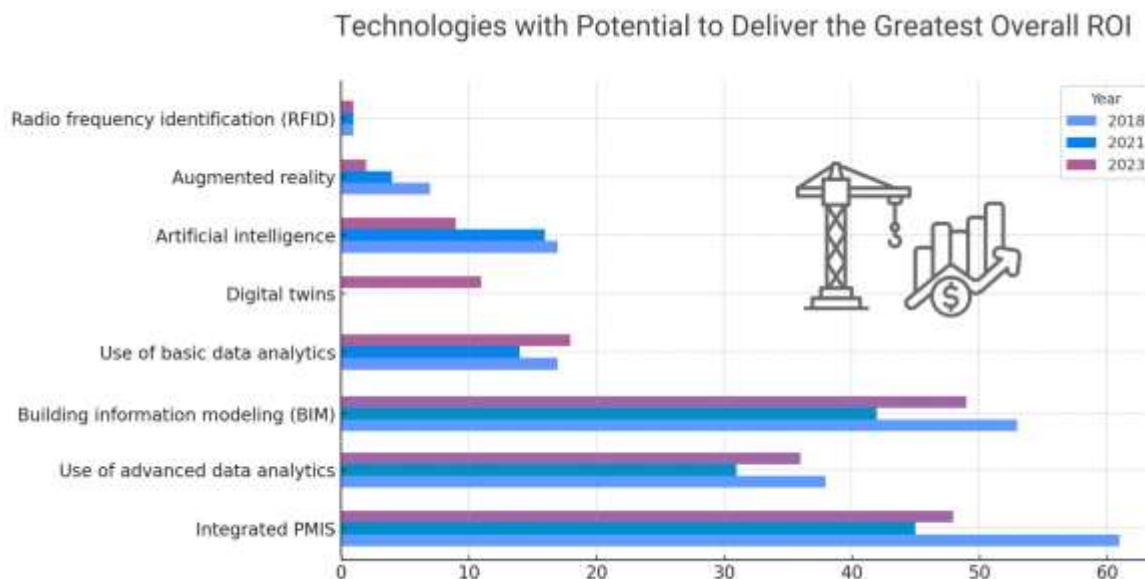
Фиг. 2 Клиентът не се интересува от прекомерния човешки фактор по пътя към реализацията на своя проект.

Въпреки всички предимства на новите инструменти и концепции, строителният сектор изостава значително от другите сектори на икономиката по отношение на внедряването на нови технологии.

Според доклада IT Metrics Key Data 2017 строителната индустрия е на последно място по разходи за ИТ сред 19 други индустрии [1].

Бързото нарастване на обема на данните и сложността на процесите се превръща в главоболие за ръководството на компанията, а основният проблем при използването на новите технологии е, че въпреки изобилието си данните остават фрагментирани, неструктурирани и често несъвместими между различните системи и софтуерни продукти. Ето защо много компании в строителния сектор сега са загрижени преди всичко за проблемите с качеството на данните, които могат да бъдат решени само с внедряването на ефективни, автоматизирани системи за управление и анализ.

Според проучване, проведено от KPMG® сред строителни мениджъри през 2023 г. [2], информационните системи за управление на проекти (PMIS), усъвършенстваният и основният анализ на данни и информационното моделиране на сгради (BIM) имат най-голям потенциал да подобрят възвръщаемостта на инвестициите в проекти (фигура 3).



**Фиг. 3 Проучване сред мениджъри на строителни компании: кои технологии ще осигурят най-висока възвръщаемост на инвестициите (ROI) в капиталови проекти? (въз основа на материали [2]).**

Решението на предизвикателствата, свързани с интегрирането на данните в бизнес процесите, е да се осигури високо качество на информацията, да се използват подходящи формати на данните и да се прилагат ефективни методи за създаване, съхранение, анализ и обработка на данните.

Осъзнаването на стойността на данните принуждава различни индустрии да се откажат от изолираните приложения и сложните бюрократични структури за управление. Вместо това фокусът се измества върху създаването на нови подходи към информационната архитектура, превъртайки компаниите в модерни предприятия, управявани от данни. Рано или късно и самата строителна индустрия ще предприеме тази стъпка, преминавайки от постепенна цифрова еволюция към истинска цифрова революция, засягаща всички компании.

Преходът към бизнес процеси, базирани на данни, няма да е лесен. Много компании ще се сблъскат с предизвикателства, тъй като ръководителите не винаги разбират как да използват хаотичните масиви от данни за подобряване на ефективността и растежа на бизнеса.

Тази книга навлиза в света на данните, където информацията се превръща в ключов стратегически ресурс, определящ ефективността и устойчивостта на бизнес процесите. С бързото нарастване на информацията компаниите са изправени пред нови предизвикателства. Цифровата трансформация вече не е просто модна дума - тя се превръща в необходимост.



**Фиг. 4** Данните и процесите са основата на строителството.

Да разбереш трансформацията означава да можеш да обясниш сложното с прости думи. Ето защо книгата е написана на достъпен език и е придружена от илюстрации на автора, създадени специално за ясно обяснение на ключови понятия. Тези схеми, диаграми и визуализации са предназначени да премахнат бариерите пред възприятието и да направят материала разбираем дори за тези, които преди са смятали подобни теми за твърде сложни. Всички илюстрации, диаграми и графики в тази книга са създадени от автора и са предназначени специално за визуализиране на ключовите понятия, описани в текста.

*Една снимка струва хиляда думи [3].*

- Фред Р. Барнард, английски илюстратор, 1927 г.

За да свържем теорията с практиката, ще използваме инструменти за изкуствен интелект (по-специално езикови модели), които ви позволяват да разработвате решения без нуждата от задълбочени познания по програмиране. Ако сте ориентирани към практическия материал и ви интересува повече практическата работа с данни, можете да пропуснете първата въвеждаща част и да преминете направо към втората част на книгата, където започва описание на конкретни примери и казуси.

Въпреки това не възлагайте прекомерни надежди на инструментите за изкуствен интелект, машинно обучение и LLM (Large Language Models) като цяло. Без качествени входни данни и задълбочено разбиране на темата дори и най-усъвършенстваните алгоритми не могат да осигурят надеждни и смислени резултати.

Главният изпълнителен директор на Microsoft Сатя Надела предупреждава за риска от балон при изкуствения интелект в началото на 2025 г. [4], като сравнява сегашния хайп с дотком балона. Той подчертава, че твърденията за достигане на етапни цели на AGI (Artificial General Intelligence) без подходяща обосновка са "безсмислена манипулация с показатели". Надела вярва, че истинският успех на ИИ трябва да се измерва с приноса му за растежа на световния БВП, а не с прекаленото наблягане на гръмки думи.

Зад всички гръмки думи за нови технологии и концепции се крие сложна и упорита работа за осигуряване на качеството на данните, параметризиране на бизнес процесите и адаптиране на инструментите към реалните задачи.

Подходът, базиран на данни, не е продукт, който можете просто да изтеглите или да купите. Той е стратегия, която трябва да се изгради. Тя започва с нов поглед върху съществуващите процеси и проблеми, а след това изисква дисциплинирано движение в избраната посока.

Водещите разработчици на софтуер и доставчици на приложения няма да бъдат двигател на промяната в строителния бранш, тъй като за много от тях подходът, основан на данни, е заплаха за установения им бизнес модел.

*Други индустрии [за разлика от строителството], като например автомобилната, вече са претърпели радикална и разрушителна промяна и тяхната цифрова трансформация е в ход. Строителните компании трябва да действат бързо и решително: пъргавите компании ще извлекат огромни ползи, докато за тези, които се колебаят, рисковете ще бъдат сериозни. Помислете за сътресенията, които цифровата фотография предизвика в този отрасъл [5].*

- Доклад на Световния икономически форум "Формиране на бъдещето на строителството", 2016 г.

Компаниите, които своевременно разпознаят възможностите и ползите от новия подход, ще получат устойчиво конкурентно предимство и ще могат да се развиват и растат, без да зависят от решенията на големите доставчици.

Това е вашият шанс не само да се справите с настъпващата бура на дигитализацията на информациите, но и да поемете контрола върху нея. В тази книга ще намерите не само анализ на настоящото състояние на отрасъла, но и конкретни препоръки за преосмисляне и преструктуриране на процесите и бизнеса ви, за да станете лидер в новата ера на строителството и да подобрите професионалния си опит.

Цифровото бъдеще на строителството не е свързано само с използването на нови технологии и програми, а с фундаментално преосмисляне на обработката на данни и бизнес моделите.

Готова ли е вашата компания за тази стратегическа промяна?



# СЪДЪРЖАНИЕ

<b>ВЪВЕДЕНИЕ .....</b>	<b>1</b>
<b>СЪДЪРЖАНИЕ .....</b>	<b>1</b>
<b>II ЧАСТ     ОТ ГЛИНЕНИТЕ ПЛОЧКИ ДО ЦИФРОВАТА РЕВОЛЮЦИЯ: КАК СЕ Е РАЗВИЛА ИНФОРМАЦИЯТА В СТРОИТЕЛСТВО .....</b>	<b>2</b>
ГЛАВА 2.1.   ЕВОЛЮЦИЯ НА ИЗПОЛЗВАНЕТО НА ДАННИ В СТРОИТЕЛНАТА ИНДУСТРИЯ .....	3
Раждането на ерата на данните в строителството .....	3
От глината и папируса до цифровите технологии .....	4
Процесът като инструмент за преживяване, базирано на данни .....	5
Цифровизация на информацията за строителния процес .....	8
ГЛАВА 2.2.   ТЕХНОЛОГИИ И СИСТЕМИ ЗА УПРАВЛЕНИЕ В СЪВРЕМЕННОТО СТРОИТЕЛСТВО .....	11
Цифровата революция и появата на модулни MRP/ERP -системи .....	11
Системи за управление на данни: от извличане на данни до бизнес предизвикателства .....	13
Корпоративен мицел: как данните се свързват с бизнес процесите .....	17
ГЛАВА 2.3.   ЦИФРОВАТА РЕВОЛЮЦИЯ И ЕКСПЛОЗИЯТА НА ДАННИ .....	20
Началото на бума на данните като еволюционна вълна .....	20
Обемът на данните, генериирани в една съвременна компания .....	22
Разходите за съхранение на данни: икономически аспект .....	23
Граници на натрупването на данни: от маса към значение .....	25
Следващи стъпки: от теория на данните към практическа промяна .....	27
<b>III ЧАСТ   КАК СТРОИТЕЛНИЯТ БИЗНЕС ПОТЪВА В ХАОС ОТ ДАННИ .....</b>	<b>29</b>
ГЛАВА 3.1.   ФРАГМЕНТАЦИЯ НА ДАННИТЕ И "СИЛОЗИ" .....	30
Колкото повече инструменти, толкова по-ефективен е бизнесът? .....	30
Силозите от данни и тяхното въздействие върху ефективността на компанията .....	32
Дублиране и липса на качество на данните като последица от разединението .....	36
HiPPO или опасността от мнения при вземането на решения .....	37
Непрекъснато нарастване на сложността и динамиката на бизнес процесите .....	40
Четвъртата индустриска революция (Индустрия 4.0) и Петата индустриска революция (Индустрия 5.0) в строителството .....	43
ГЛАВА 3.2.   ПРЕВРЪЩАНЕ НА ХАОСА В РЕД И НАМАЛЯВАНЕ НА СЛОЖНОСТТА .....	46
Излишният код и затворените системи като пречка за подобряване на производителността ..	46
От отделни системи към единен склад за данни .....	48

Интегрираните системи за съхранение позволяват преминаването към агенти с изкуствен интелект .....	50
От събиране на данни до вземане на решения: пътят към автоматизацията .....	53
Следващи стъпки: превръщане на хаоса в управляема система .....	55
<b>IV ЧАСТ РАМКА ЗА ДАННИ В БИЗНЕС ПРОЦЕСИТЕ В СТРОИТЕЛСТВОТО.....</b>	<b>57</b>
ГЛАВА 4.1. ТИПОВЕ ДАННИ В СТРОИТЕЛСТВОТО .....	58
Най-важните типове данни в строителната индустрия .....	58
Структурирани данни .....	62
Релационни бази данни RDBMS и език за заявки SQL .....	64
SQL - заявки в базите данни и нови тенденции .....	66
Неструктурирани данни .....	68
Текстови данни: между неструктурен хаос и структуриран хаос .....	69
Полуструктурни и слабо структурирани данни .....	70
Геометрични данни и тяхното приложение .....	71
CAD данни: от проектирането до съхранението на данни .....	74
Появата на концепцията BIM (BOM) и използването на CAD в процесите .....	78
ГЛАВА 4.2. ХАРМОНИЗИРАНЕ И СТРУКТУРИРАНЕ НА ДАННИ .....	85
Запълване на системите с данни в строителната индустрия .....	85
Трансформация на данните: критичната основа на съвременния бизнес анализ .....	88
Модели на данни: връзки в данните и връзки между елементите .....	91
Запазени формати и тяхното въздействие върху цифровите процеси .....	96
Отворените формати променят подхода към цифровизацията .....	100
Смяна на парадигмата: отвореният код като край на ератата на господство на доставчиците на софтуер .....	101
Структурирани отворени данни: основата на цифровата трансформация .....	104
ГЛАВА 4.3. LLM И ТЯХНАТА РОЛЯ В ОБРАБОТКАТА НА ДАННИ И БИЗНЕС ПРОЦЕСИТЕ .....	107
Чат стаи на LLM: ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok за автоматизиране на процесите на данни .....	107
Големи езикови модели LLM: как работи .....	108
Използване на местни LLM за чувствителни фирмени данни .....	111
Пълен контрол върху AI в компанията и как да внедрите свой собствен LLM .....	113
RAG: Интелигентен LLM - помощници с достъп до корпоративни данни .....	115
ГЛАВА 4.4. IDE С ПОДДРЪЖКА НА LLM И БЪДЕЩИ ПРОГРАМНИ ПРОМЕНИ .....	118
Избор на IDE: от експерименти по LLM до бизнес решения .....	118
IDE с поддръжка на LLM и бъдещи промени в програмирането .....	120

Python Pandas: незаменим инструмент за работа с данни .....	121
DataFrame: универсален формат за таблични данни .....	126
Следващи стъпки: изграждане на устойчива рамка за данни .....	129
<b>V ЧАСТ КАЧЕСТВО НА ДАННИТЕ: ОРГАНИЗАЦИЯ, СТРУКТУРИРАНЕ, МОДЕЛИРАНЕ .....</b>	<b>131</b>
ГЛАВА 5.1. ПРЕОБРАЗУВАНЕ НА ДАННИ В СТРУКТУРИРАНА ФОРМА .....	132
Научете се как да превръщате документи, PDF, снимки и текстове в структурирани формати .....	132
Пример за преобразуване на PDF -документ в таблица.....	133
Конвертиране на JPEG, PNG изображение в структурирана форма .....	137
Конвертиране на текстови данни в структурирана форма.....	140
Конвертиране на CAD данни (BIM) в структуриран вид.....	143
Доставчиците на CAD решения се насочват към структурирани данни .....	148
ГЛАВА 5.2. КЛАСИФИКАЦИЯ И ИНТЕГРАЦИЯ: ОБЩ ЕЗИК ЗА СТРОИТЕЛНИТЕ ДАННИ.....	151
Бързината на вземане на решения зависи от качеството на данните .....	151
Стандартизация и интеграция на данни .....	152
Цифровата оперативна съвместимост започва с изискванията.....	155
Общ език за конструиране: ролята на класifikаторите в цифровата трансформация.....	158
Masterformat, OmniClass, Uniclass и CoClass: еволюция на системите за класификация.....	161
ГЛАВА 5.3. МОДЕЛИРАНЕ НА ДАННИ И ЦЕНТЪР ЗА ВЪРХОВИ ПОСТИЖЕНИЯ .....	166
Моделиране на данни: концептуален, логически и физически модел .....	166
Практическо моделиране на данни в контекста на строителството .....	170
Създаване на база данни с помощта на LLM .....	172
Центрър за върхови постижения (ЦВП) за моделиране на данни.....	174
ГЛАВА 5.4. СИСТЕМАТИЗИРАНЕ НА ИЗИСКВАНИЯТА И ВАЛИДИРАНЕ НА ИНФОРМАЦИЯТА .....	178
Събиране и анализ на изисквания: трансформиране на комуникациите в структурирани данни .....	178
Блок-схеми на процеси и ефективност на концептуалните рамки.....	182
Структурирани изисквания и регулярни изрази RegEx .....	184
Събиране на данни за процеса на проверка .....	189
Проверка на данните от и резултати от проверката .....	191
Визуализация на резултатите от проверката на .....	196
Сравнение на проверките на качеството на данните с нуждите на човешкия живот .....	198
Следващи стъпки: превръщане на данните в точни изчисления и планове .....	200
<b>VI ЧАСТ ИЗЧИСЛЯВАНЕ НА РАЗХОДИТЕ И ВРЕМЕТО: ВКЛЮЧВАНЕ НА ДАННИ В СТРОИТЕЛНИТЕ ПРОЦЕСИ .....</b>	<b>202</b>

---

ГЛАВА 6.1. ИЗЧИСЛЯВАНЕ НА РАЗХОДИ И ОЦЕНКИ ЗА СТРОИТЕЛНИ ПРОЕКТИ .....	203
Основи на строителството: оценяване на количеството, разходите и времето .....	203
Методи за изчисляване на прогнозната стойност на проектите .....	205
Метод, базиран на ресурсите остойностяване и оценяване в строителството .....	205
База данни за строителни ресурси: каталог на строителни материали и работи .....	206
Изготвяне на изчисления и остойностяване на работите въз основа на ресурсна база .....	208
Окончателно остойностяване на проекта: от оценки до бюджети .....	213
ГЛАВА 6.2. ИЗЧИСЛЯВАНЕ НА КОЛИЧЕСТВАТА И АВТОМАТИЧНО СЪЗДАВАНЕ НА ОЦЕНКИ И ГРАФИЦИ .....	217
Преминаване от 3D към 4D и 5D: използване на обемни и количествени параметри .....	217
5D атрибути и получаване на обеми на атрибути от CAD .....	217
QTO Quantity Take-Off: групиране на данни за проекта по атрибути .....	221
Автоматизация на QTO с помощта на LLM и структурирани данни .....	226
QTO изчисление на целия проект с помощта на групови правила от електронна таблица на Excel .....	230
ГЛАВА 6.3. 4Г, 6Г -8Г И ИЗЧИСЛЯВАНЕ НА ЕМИСИИТЕ НА ВЪГЛЕРОДЕН ДИОКСИД CO <sub>2</sub> .....	236
4D модел: интегриране на времето в строителните оценки .....	236
График на строителството и неговото автоматизиране въз основа на данни за разходите .....	237
Разширени атрибутни слоеве 6D -8D: от енергийна ефективност до осигуряване на безопасност .....	240
Оценка на емисиите на CO <sub>2</sub> и изчисляване на емисиите на въглероден диоксид от строителни проекти .....	243
ГЛАВА 6.4. СИСТЕМИ ЗА СТРОИТЕЛСТВО ERP И PMIS .....	248
Строителни ERP -системи на примера на изчисления и оценки .....	248
PMIS: междуинно звено между ERP и строителната площадка .....	254
Спекулации, печалба, изолираност и липса на прозрачност в ERP и PMIS .....	255
Краят на ерата на затворените ERP /PMIS: строителната индустрия се нуждае от нови подходи .....	258
Следващи стъпки: ефективно използване на данните от проекта .....	260
<b>VII ЧАСТ CAD И BIM: МАРКЕТИНГ, РЕАЛНОСТ И БЪДЕЩЕТО НА ДАННИТЕ ЗА ПРОЕКТИРАНЕ В СТРОИТЕЛСТВО.....</b>	<b>264</b>
ГЛАВА 7.1. ПОЯВА НА КОНЦЕПЦИИТЕ НА BIM- В СТРОИТЕЛНАТА ИНДУСТРИЯ .....	265
История на появата на BIM и отворен BIM като маркетингови концепции на доставчиците на CAD- .....	265
Реалността на BIM: вместо интегрирани бази данни - затворени модулни системи .....	268
Появата на отворения формат IFC в строителната индустрия .....	270

Проблем с формата IFC в зависимост от геометричното ядро .....	272
Появяване в конструкцията на темата за семантиката и онтологията .....	276
Защо семантичните технологии не оправдават очакванията в строителната индустрия .....	278
<b>ГЛАВА 7.2. ЗАТВОРЕНИ ФОРМАТИ НА ПРОЕКТИТЕ И ПРОБЛЕМИ С ОПЕРАТИВНАТА СЪВМЕСТИМОСТ .....</b>	<b>281</b>
Затворени данни и спадаща производителност: задънената улица на CAD индустрията (BIM) .....	281
Митът за оперативната съвместимост между CAD системите .....	283
Към USD и гранулирани данни т.....	287
<b>ГЛАВА 7.3. ГЕОМЕТРИЯ В СТРОИТЕЛСТВОТО: ОТ ЛИНИИ ДО КУБИЧНИ МЕТРИ .....</b>	<b>291</b>
Когато линиите се превръщат в пари или защо строителите се нуждаят от геометрия.....	291
От линии към обеми: как площта и обемът се превръщат в данни .....	291
Преминаване към MESH, USD и полигони: използване на теселиране за геометрия .....	294
LOD, LOI, LOMD - уникална класификация на детайлите в CAD (BIM).....	295
Нови CAD стандарти (BIM) - AIA, BEP, IDS, LOD, COBie .....	298
<b>ГЛАВА 7.4. ПАРАМЕТРИЗИРАНЕ НА ДИЗАЙНА И ИЗПОЛЗВАНЕ НА LLM ЗА РАБОТА С CAD .....</b>	<b>303</b>
Илюзията за уникалност на CAD данните (BIM): пътят към анализи и отворени формати .....	303
Проектиране чрез параметри: бъдещето на CAD и BIM .....	306
Възникване на LLM в процесите на обработка на данни в CAD проектирането .....	309
Автоматизиран анализ на DWG -файлове с LLM и Pandas .....	312
Следващи стъпки: преминаване от затворени формати към отворени данни .....	318
<b>VIII ЧАСТ ВЗЕМАНЕ НА РЕШЕНИЯ, БАЗИРАНИ НА ДАННИ, АНАЛИЗ, АВТОМАТИЗАЦИЯ И МАШИННО ОБУЧЕНИЕ .....</b>	<b>321</b>
<b>ГЛАВА 8.1. АНАЛИЗ НА ДАННИ И ВЗЕМАНЕ НА РЕШЕНИЯ, ОСНОВАНИ НА ДАННИ .....</b>	<b>322</b>
Данните като ресурс при вземането на решения.....	322
Визуализиране на данни от: ключ към разбирането и вземането на решения .....	326
Ключови показатели за ефективност и възвръщаемост на инвестициите .....	328
Информационни табла и табла за управление: визуализиране на показатели за ефективно управление .....	330
Анализ на данни и изкуството да се задават въпроси .....	331
<b>ГЛАВА 8.2. ПОТОК НА ДАННИ БЕЗ РЪЧНИ УСИЛИЯ: ЗАЩО Е НЕОБХОДИМ ETL .....</b>	<b>334</b>
Автоматизация на ETL: намаляване на разходите и ускоряване на обработката на данни.....	334
ETL Извличение: събиране на данни .....	338
ETL Трансформиране: прилагане на правила за валидиране и трансформиране .....	341

---

ETL Load: Визуализирайте резултатите от в диаграми и графики .....	343
ETL Зареждане: Автоматично създаване на PDF документи .....	349
ETL Load: автоматично генериране на документи от FPDF .....	350
ETL Натоварване: Отчитане и зареждане на други системи.....	354
ETL с LLM: Визуализиране на данни от PDF -документи .....	355
<b>ГЛАВА 8.3. АВТОМАТИЧЕН КОНВЕЙЕР ЗА ETL (ТРЪБОПРОВОД)</b> .....	<b>361</b>
Тръбопровод: Автоматичен ETL конвейер данни.....	361
Тръбопровод -ETL процес на валидиране на данни с LLM .....	365
Тръбопровод -ETL: проверка на данните и информацията за елементите на проекта в CAD (BIM) .....	367
<b>ГЛАВА 8.4. ОРКЕСТРАЦИЯ НА ETL И РАБОТНИ ПОТОЦИ: ПРАКТИЧЕСКИ РЕШЕНИЯ.....</b>	<b>373</b>
DAG и Apache Airflow: автоматизация на работния процес и оркестрация .....	373
Apache Airflow: практическо приложение автоматизация на ETL.....	374
Apache NiFi за маршрутизиране и преобразуване на данни.....	378
n8n Low-Code, No-Code оркестрация на процеси.....	379
Следващи стъпки: преминаване от ръчни операции към аналитични решения .....	382
<b>IX ЧАСТ СЪХРАНЕНИЕ И УПРАВЛЕНИЕ НА ДАННИ В СТРОИТЕЛСТВОТО .....</b>	<b>384</b>
<b>ГЛАВА 9.1. ИНФРАСТРУКТУРА ЗА ДАННИ: ОТ ФОРМАТИ ЗА СЪХРАНЕНИЕ ДО ЦИФРОВИ ХРАНИЛИЩА .....</b>	<b>385</b>
Атомите на данни: основата на ефективното управление на информацията .....	385
Съхранение на информация: файлове или данни .....	386
Съхраняване на големи обеми данни: анализ на популярните формати и тяхната ефективност .....	388
Оптимизиране на съхранението с Apache Parquet.....	391
DWH: Data Warehouse складове за данни .....	394
Езеро от данни - еволюция на ETL към ELT: от традиционно почистване към гъвкава обработка .....	395
Архитектура на езерото за данни: синергия на складове и езера за данни.....	397
CDE, PMIS, ERP или DWH и езеро от данни .....	400
<b>ГЛАВА 9.2. УПРАВЛЕНИЕ НА ХРАНИЛИЩА ЗА ДАННИ И ПРЕДОТВРАТИВАНЕ НА ХАОС .....</b>	<b>403</b>
Векторни бази данни и гранична кутия.....	403
Управление на данните), минимализъм на данните) и блато на данните) .....	406
DataOps и VectorOps: нови стандарти за данни .....	409
Следващи стъпки: от хаотично съхранение към структурирано съхранение .....	411
<b>X ЧАСТ ГОЛЕМИ ДАННИ, МАШИННО ОБУЧЕНИЕ И ПРОГНОЗИ .....</b>	<b>413</b>

<b>ГЛАВА 10.1. ГОЛЕМИ ДАННИ И ТЕХНИЯТ АНАЛИЗ .....</b>	<b>414</b>
Големи данни в строителството: от интуиция към предвидимост .....	414
Въпросът за осъществимостта на големите данни: корелация, статистика и извадка от данни .....	415
Големи данни: анализ на данни от набора от данни за милиони разрешителни за строеж в Сан Франциско .....	418
Пример за големи данни, базирани на CAD данни (BIM) .....	424
IoT Интернет на нещата и интелигентни договори .....	428
<b>ГЛАВА 10.2. МАШИННО ОБУЧЕНИЕ И ПРОГНОЗИ .....</b>	<b>432</b>
Машинното обучение и изкуственият интелект ще променят начина, по който създаваме ....	432
От субективна оценка към статистическа прогноза .....	435
Набор от данни "Титаник": "Здравей, свят" в света на анализите данни и големи данни .....	436
Машинно обучение в действие: от пътниците на "Титаник" до управлението на проекти .....	442
Предвиждания и прогнози въз основа на исторически данни .....	446
Ключови концепции за машинно обучение .....	448
<b>ГЛАВА 10.3. ПРОГНОЗИРАНЕ НА РАЗХОДИ И СРОКОВЕ С ПОМОЩТА НА МАШИННО ОБУЧЕНИЕ</b>	<b>451</b>
Пример за използване на машинно обучение за намиране на разходите и сроковете на проекта .....	451
Прогнозиране на разходите и времето за изпълнение на проекта чрез линейна регресия ....	453
Прогнозиране на разходите и времето за изпълнение на проекта с помощта на алгоритъма на най-близките съседи (k-NN) .....	456
Следващи стъпки: от съхранение до анализ и прогнозиране .....	460
<b>I ЧАСТ СТРОИТЕЛНАТА ИНДУСТРИЯ В ЦИФРОВАТА ЕРА. ВЪЗМОЖНОСТИ И ПРЕДИЗВИКАТЕЛСТВА.....</b>	<b>463</b>
<b>ГЛАВА 11.1. СТРАТЕГИИ ЗА ОЦЕЛЯВАНЕ: ИЗГРАЖДАНЕ НА КОНКУРЕНТНИ ПРЕДИМСТВА .....</b>	<b>464</b>
Корелации вместо изчисления: бъдещето на строителните анализи .....	464
Подход, основан на данни, в строителството: ново ниво на инфраструктурата .....	468
Цифровият офис от следващо поколение: как AI променя работното пространство .....	470
Отворените данни и Uberisation са заплаха за съществуващия строителен бизнес .....	473
Неразрешените проблеми на уберализацията като последен шанс да се използва времето за трансформация .....	476
<b>ГЛАВА 11.2. ПРАКТИЧЕСКО РЪКОВОДСТВО ЗА ПРИЛАГАНЕ НА ПОДХОД, ОСНОВАН НА ДАННИ</b>	<b>481</b>
От теория към практика: пътна карта за цифрова трансформация в строителството .....	481
Полагане на цифрова основа: 1-5 стъпки към цифрова зрълост .....	483
Отключване на потенциала на данните: 5-10 стъпки към цифровата зрълост .....	488

Пътна карта за трансформация: от хаос към компания, управлявана от данни .....	495
Строителство в Индустринг 5.0: Как да печелим, когато вече не можем да се крием.....	498
<b>ЗАКЛЮЧЕНИЕ.....</b>	<b>500</b>
<b>ЗА АВТОРА .....</b>	<b>503</b>
<b>ПО-НАТАТЬШНА ВРЪЗКА .....</b>	<b>504</b>
<b>КОМЕНТАР НА ПРЕВОДА .....</b>	<b>504</b>
<b>ДРУГИ УМЕНИЯ И КОНЦЕПЦИИ .....</b>	<b>505</b>
<b>РЕЧНИК .....</b>	<b>509</b>
<b>СПИСЪК С РЕФЕРЕНЦИИ И ОНЛАЙН МАТЕРИАЛИ.....</b>	<b>516</b>
<b>ИНДЕКС НА ПРЕДМЕТИТЕ .....</b>	<b>533</b>

## МАКСИМАЛНО УДОБСТВО С ПЕЧАТНАТА ВЕРСИЯ

Вие притежавате бесплатна цифрова версия на **Data-Driven Construction**. За по-удобна работа и бърз достъп до материалите ви препоръчваме да обърнете внимание на **печатното издание**:



**■ Винаги под ръка:** книгата в печатен формат ще се превърне в надежден работен инструмент, който ще ви позволи бързо да намерите и използвате правилните визуализации и диаграми във всяка работна ситуация.

**■ Висококачествени илюстрации:** всички изображения и графики в печатното издание са представени с максимално качество.

**■ Бърз достъп до информация:** лесна навигация, възможност за правене на бележки, отметки и работа с книгата навсякъде.

Закупувайки пълната печатна версия на книгата, вие получавате удобен инструмент за комфортна и ефективна работа с информация: възможност за бързо използване на визуални материали в ежедневните задачи, бързо намиране на необходимите схеми и водене на бележки. Освен това с покупката си подкрепяте разпространението на отвореното знание.

Поръчайте печатна версия на книгата на адрес: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



# I ЧАСТ

## ОТ ГЛИНЕНИТЕ ПЛОЧКИ ДО ЦИФРОВАТА РЕВОЛЮЦИЯ: КАК СЕ Е РАЗВИЛА ИНФОРМАЦИЯТА В СТРОИТЕЛСТВОТО

В първата част на книгата се разглежда историческата еволюция на управлението на данни в строителната индустрия - от примитивните записи на физически носители до съвременните цифрови екосистеми. Анализира се трансформацията на технологиите за управление на информацията, появата на ERP-системите и въздействието на фрагментацията на данните върху ефективността на бизнес процесите. Особено внимание се обръща на процеса на дигитализация на информацията и на нарастващото значение на обективния анализ вместо субективната експертна оценка. Експоненциалното нарастване на обема на информацията, с което се сблъсква съвременната строителна индустрия, и свързаните с това предизвикателства пред корпоративните системи са разгледани подробно. Изследвано е позиционирането на строителната индустрия в контекста на четвъртата и петата индустриална революция, както и потенциалът за използване на изкуствен интелект и ориентирани към данните подходи за създаване на устойчиво конкурентно предимство.

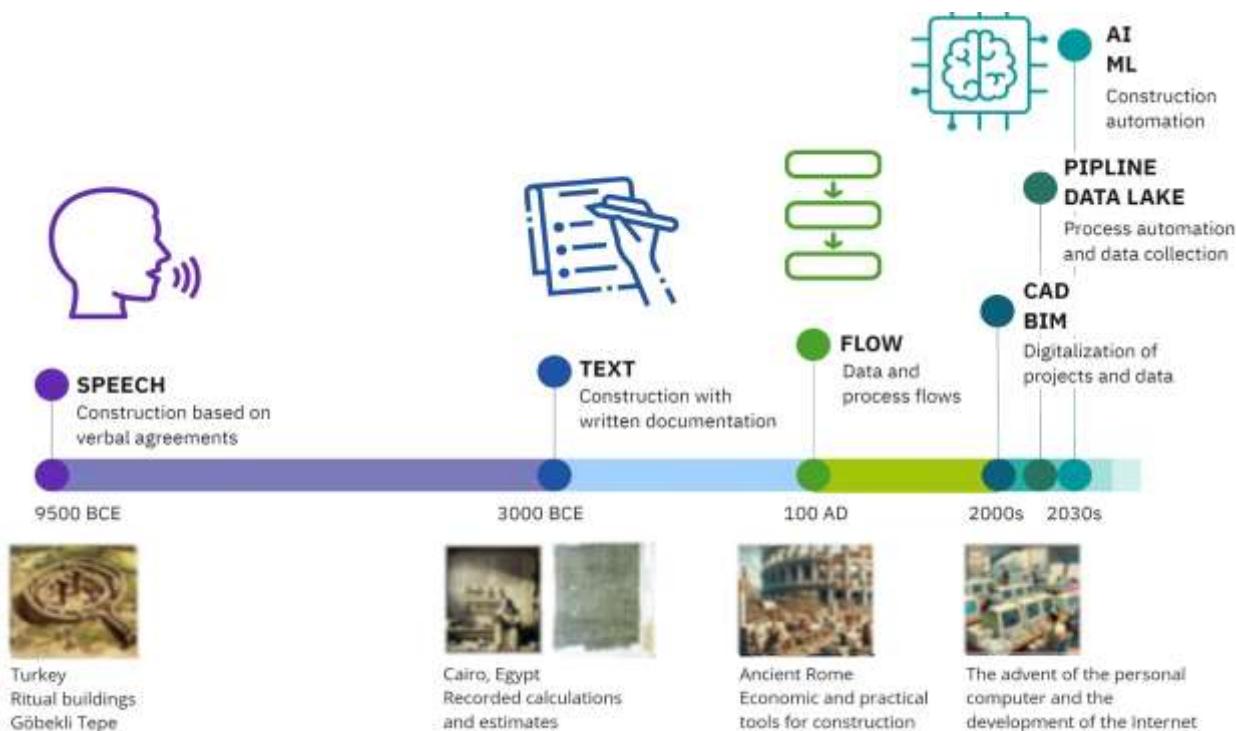
## ГЛАВА 1.1.

# ЕВОЛЮЦИЯ НА ИЗПОЛЗВАНЕТО НА ДАННИ В СТРОИТЕЛНАТА ИНДУСТРИЯ

### Раждането на ерата на данните в строителството

Преди около 10 000 години, през епохата на неолита, човечеството извършва революционен преход в своето развитие, като изоставя номадския начин на живот в полза на уседналия, което води до появата на първите примитивни сгради, направени от глина, дърво и камък [6]. От този момент започва историята на строителната индустрия.

С развитието на цивилизациите архитектурата става все по-сложна, което води до появата на първите ритуални храмове и обществени сгради. Нарастващата сложност на архитектурните проекти налага на инженерите и мениджърите от древността да създадат първите записи и изчисления. Първите записи върху глинени плочки и папируси често са включвали описание на логиката, стояща зад изчисляването на количеството необходими строителни материали, тяхната цена и изчисляването на заплащането за извършената работа [7]. Така започва ерата на използване на данни в строителството - много преди появата на съвременните цифрови технологии (фигура 1.1-1).

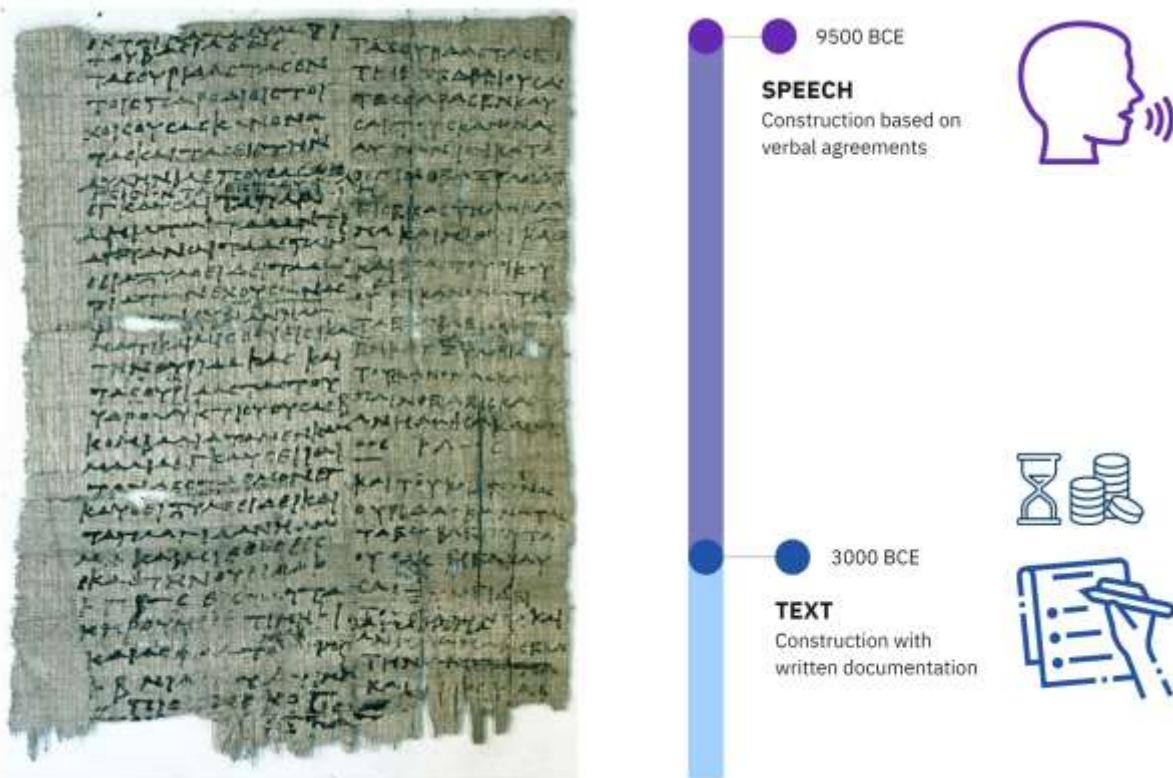


Фиг. 1.1-1 Хронология на развитието на информационните технологии в строителството: от вербалната информация до изкуствения интелект.

## От глината и папируса до цифровите технологии

Първите документални свидетелства за строителството датират от периода на строителството на пирамиди, около 3000-4000 г. пр.н.е.[7]. Оттогава насам воденето на писмени документи улеснява и съпътства напредъка в строителството, позволявайки натрупването и систематизирането на знания, които през следващите 10 000 години водят до значителни нововъведения в строителните методи и архитектурата.

Използването на първите физически носители в строителството, като глинени плочки, папируси отпреди хиляди години (фигура 1.1-2) или хартия "A0" през 80-те години на миналия век, за записване на данни първоначално не е било предназначено за прилагане на тази информация в нови проекти. Основната цел на такива записи е била да се опише подробно текущото състояние на проекта, включително изчисления на необходимите материали и стойността на работата. По подобен начин в днешния свят наличието на цифрови проектни данни и модели невинаги гарантира прилагането им в бъдещи проекти и често служи главно като информация за текущи изчисления на необходимите материали и разходите за строителство.



Фиг. 1.1-2 Папирус от III в. пр.н.е., в който се описват разходите за боядисване на различни видове прозорци в кралски дворец с помощта на техниката енкаустика.

На човечеството са били необходими около 5000 години, за да премине от устни разговори към писмени документи в управлението на строителни проекти, и също толкова време, за да премине от хартия към цифрови данни като основен ресурс за планиране и контрол.

Точно както развитието на търговията и паричните отношения стимулира появата на писмеността и първите адвокати за решаване на спорове, така и първите записи на материалните разходи и обемите на работа в строителството водят до появата на първите мениджъри в строителния бранш, чито задължения включват документиране, наблюдение и отговорност за ключова информация за сроковете и разходите по проекта.

Днес данните играят много по-важна роля: те не само регистрират взетите решения, но и се превръщат в инструмент за прогнозиране и моделиране на бъдещето. Това е основата, върху която се гради съвременният процесен подход в управлението на проекти - превръщането на натрупания опит в система за вземане на решения, основана на структурирани и проверими данни.

### Процесът като инструмент за преживяване, базирано на данни

В основата на всеки процес е превръщането на миналия опит в инструмент за планиране на бъдещето. Опитът в съвременния смисъл на думата е структуриран набор от данни, които могат да бъдат анализирани, за да се направят информирани прогнози.

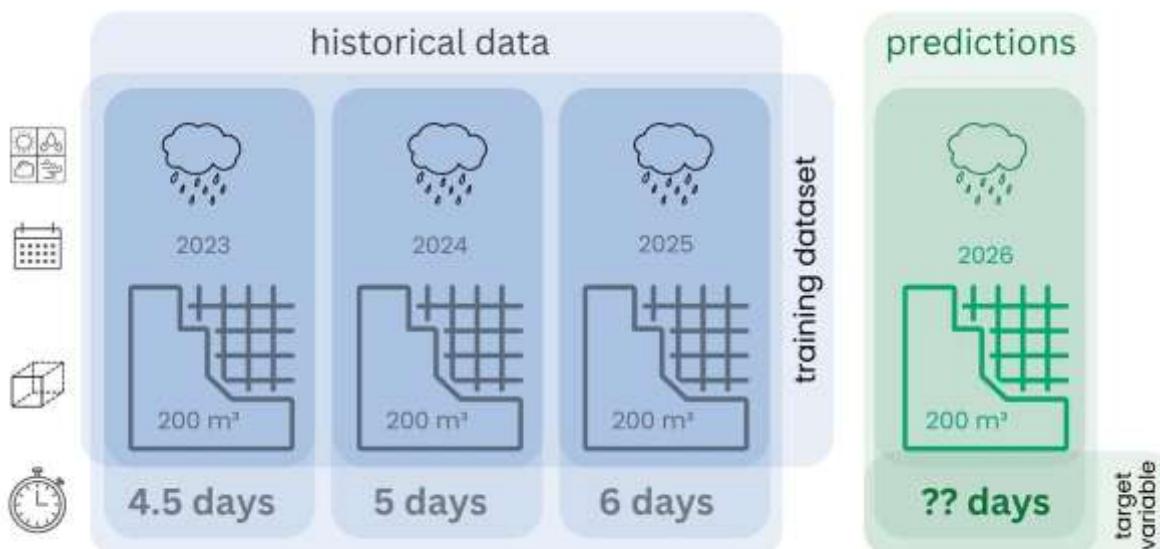
Историческите данни са в основата на прогнозирането, тъй като те ясно показват резултатите от извършената работа и дават представа за факторите, които влияят на тези резултати.

Нека вземем конкретен пример от монолитното строителство: обикновено при планирането на сроковете на работите се вземат предвид обемът на бетона, сложността на конструкцията и метеорологичните условия. Да предположим, че конкретен бригадир на обекта или историческите данни на компанията за последните три години (2023-2025 г.) показват, че изливането на монолитна конструкция с площ 200 m<sup>2</sup> в дъждовно време отнема между 4,5 и 6 дни (фиг. 1.1-3). Именно тези натрупани статистически данни се превръщат в основа за прогнозиране на времето за изпълнение и остойностяване на ресурсите при планиране на подобна работа в бъдещи проекти. Въз основа на тези исторически данни бригадирът или оценителят може да направи информирана прогноза, основана на опита, за времето, необходимо за завършване на бъдещи подобни работи през 2026 г. при сходни условия.

В случая с оценките на времето аналитичният процес действа като механизъм за трансформиране на разнородни данни в структуриран опит и след това в инструмент за прецизно планиране. Данните и процесите са единна екосистема, в която едното не може да съществува без другото.

Пребройте това, което може да се преброи, измерете това, което може да се измери, и направете това, което не може да се измери, измеримо [8].

- Галилео Галилей

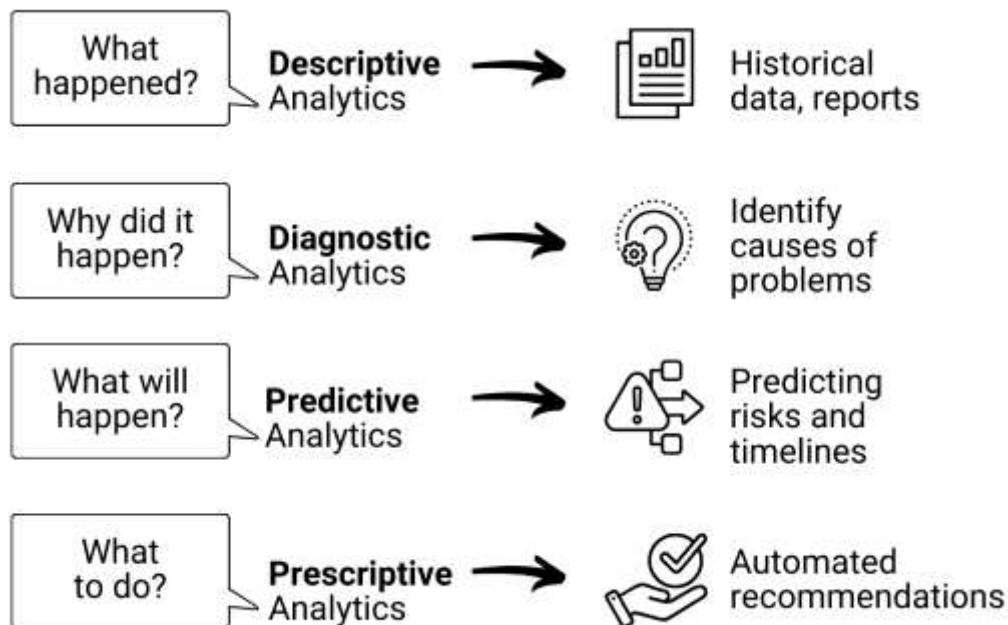


**Фиг. 1.1-3 Историческите данни служат като набор от данни за обучение за прогнозиране на една от стойностите в бъдеще**

В днешната бизнес среда анализът на данни се превръща в критичен компонент на ефективното управление на проекти, оптимизацията на процеси и вземането на стратегически решения. Строителният бранш постепенно овладява четири ключови нива на анализ, всяко от които дава отговор на конкретен въпрос и осигурява уникални ползи (фиг. 1.1-4):

- **Дескриптивен анализ** - отговаря на въпроса "какво се е случило?" и предоставя исторически данни и доклади за минали събития и резултати: през последните три години (2023-2025 г.) за изливането на монолитна конструкция с площ 200 m<sup>2</sup> в дъждовно време са били необходими между 4,5 и 6 дни.
- **Диагностичен анализ** - дава отговор на въпроса "защо се случи това?", като идентифицира причините за проблемите: анализът показва, че времето за изливане на монолитната конструкция се е увеличило поради дъждовното време, което е забавило процеса на втвърждане на бетона.
- **Прогнозен анализ** - ориентиран към бъдещето, предвиждащ възможни рискове и срокове за изпълнение чрез отговор на въпроса "какво ще се случи?": въз основа на исторически данни се предвижда, че изливането на подобна монолитна конструкция с площ 200 m<sup>2</sup> в дъждовно време през 2026 г. ще отнеме приблизително 5,5 дни, като се вземат предвид всички известни фактори и тенденции.
- **Предписващ анализ** - предоставя автоматизирани препоръки и отговаря на въпроса

"какво да се направи?", като позволява на компаниите да изберат оптималните действия: За да се оптимизира работата, например, се препоръчва: да се използват специални добавки за ускоряване на втвърдяването на бетона в условия на висока влажност; да се планира изливането за периоди с най-малка вероятност за валежи; да се организират временни навеси за конструкцията, което ще намали времето за работа до 4-4,5 дни дори при неблагоприятни метеорологични условия.



Фиг. 1.1-4 Основните видове анализи: от описание на миналото до автоматизирано вземане на решения.

Пълноценната цифрова трансформация, която предполага преход към системен анализ и управление, основано на данни, изисква не само аутсорсинг, но и сформиране на компетентен вътрешен екип. Основните членове на такъв екип трябва да бъдат продуктови мениджъри, инженери по данни, анализатори и разработчици, които ще работят в тясно сътрудничество с бизнес звената (фиг. 4.3-9). Това сътрудничество е необходимо, за да се задават интелигентни аналитични въпроси и ефективно да се параметризират задачите за вземане на бизнес решения. В информационното общество данните се превръщат не само в спомагателен инструмент, но и в основа за прогнозиране и оптимизиране.

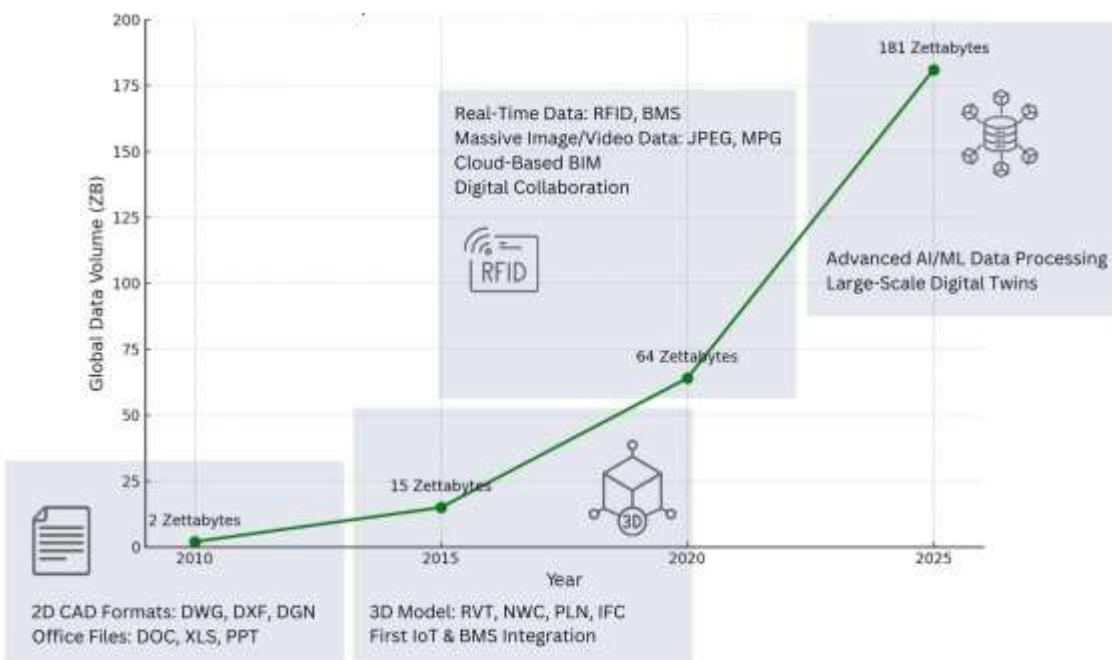
В строителството цифровата трансформация променя из основи начина, по който се проектират, управляват и експлоатират съоръженията. Този процес се нарича цифровизация на информациите - при него всички аспекти на строителния процес се цифровизират в цифрова форма, подходяща за анализ.

## Цифровизация на информацията за строителния процес

В продължение на хилядолетия обемът на информацията, записана в строителството, почти не се е променял, но през последните десетилетия той бързо нараства (фигура 1.1-5).

Според проучването на PwC® "Managed Data. От какво се нуждаят студентите, за да успеят в бързо променящия се бизнес свят" (2015 г.) [9], 90% от всички данни в света са създадени през последните две години (към 2015 г.). Повечето компании обаче не използват пълноценно тези данни, тъй като те или остават в изолирани системи, или просто се архивират без реален анализ.

Увеличаването на обема на данните се ускори през последните години, като се удвои от 15 зетабайта през 2015 г. до 181 зетабайта през 2025 г. [10]. Всеки ден сървърите на строителните и проектантските фирми се пълнят с проектна документация, работни графики, изчисления и калкулации, финансови отчети. За 2D/3D -чертежи се използват форматите DWG, DXF и DGN, а за 3D модели - RVT, NWC, PLN и IFC™. Текстовите документи, таблиците и презентациите се записват във форматите DOC, XLSX и PPT. Освен видео и изображения от строителната площадка - в MPG и JPEG, се съхраняват и данни в реално време от IoT компоненти, RFID® тагове (идентификация и проследяване) и BMS системи за управление на сгради (мониторинг и контрол)



Фиг. 1.1-5 Параболичен ръст на данните 2010-2025 г. (въз основа на [10]).

С бързото нарастване на информацията строителната индустрия се сблъскава с необходимостта не само да събира и съхранява данни, но и да гарантира тяхната проверка, валидиране, измеримост и аналитична обработка. Днес отрасът преминава през активна фаза на информационна дигитализация - систематично преобразуване на всички аспекти на строителната дейност в цифрова форма, подходяща за анализ, интерпретация и автоматизация.

Дигитализирането на информацията означава да се вземе информация за всички субекти и елементи на даден строителен проект и на самия строителен процес - включително и за тези, които досега изобщо не сме считали за информация - и да се преобразува във формат на данни, за да може информацията да се определи количествено и да се анализира лесно.

В контекста на строителството това означава събиране и цифровизиране на информация всички елементи на проектите и всички процеси - от движението на машините и хората на строителната площадка до метеорологичните и климатичните условия на строителната площадка, текущите цени на материалите и лихвените проценти на централните банки - с цел създаване на аналитични модели.

*Ако можете да измерите това, за което говорите, и да го изразите в цифри, значи знаете нещо по въпроса. Но ако не можете да го изразите количествено, знанията ви са крайно ограничени и нездадоволителни. То може да е отправна точка, но не е нивото на истинското научно познание. [11].*

- У. Томсън (lord Kelvin), 1824-1907 г., британски учен

Цифровизацията на информацията далеч надхвърля традиционния подход към събирането на информация, при който се записваха само основни показатели като човекочасове или действителни разходи за материали. Днес практически всяко събитие може да бъде превърнато в поток от данни, подходящ за задълбочен анализ с помощта на усъвършенствани инструменти за анализ и техники за машинно обучение. Строителната индустрия е претърпяла фундаментална промяна от хартиени чертежи, електронни таблици в Excel и устни инструкции към цифрови системи (фигура 1.2-4), в които всеки елемент на проекта се превръща в източник на данни. Дори на служителите - от инженерите до строителните работници на обекта - вече се гледа като на съвкупност от цифрови променливи и набори от данни.

Според доклада на KPMG "Familiar Challenges - New Approaches: Global Construction Survey 2023", цифровите близнци, изкуственият интелект (ИИ) и големите данни, се очертават като ключови фактори за подобряване на рентабилността на проектите [2].

Съвременните технологии не само опростяват събирането на информация, като го правят до голяма степен автоматично, но и радикално намаляват разходите за съхранение на данни. В резултат на това компаниите се отказват от селективния подход и предпочитат да съхраняват целия масив от информация за последващ анализ (фиг. 2.1-5), което открива потенциални възможности за оптимизиране на процесите в бъдеще.

Цифровизацията на информацията и цифровизацията дават възможност да се разкрие скритата, неизползвана досега стойност на информацията. Ако са организирани правилно, данните могат да бъдат използвани повторно, интерпретирани и интегрирани в нови услуги и решения.

В бъдеще цифровизацията на информацията вероятно ще доведе до пълна автоматизация на управлението на документите, въвеждане на самоуправляващи се строителни процеси и появата на нови професии - анализатори на данни в строителството, експерти по управление на проекти с изкуствен интелект и цифрови инженери. Строителните проекти ще се превърнат в динамични източници на информация, а вземането на решения няма да се основава на интуиция или субективен опит, а на надеждни и възпроизведими цифрови факти

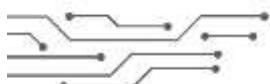
*Информацията е петролът на 21-ви век, а анализът е двигателят с вътрешно горене [12].*

- Питър Сондергаард, старши вицепрезидент, Gartner®

Според доклада IoT Analytics 2024 [13] глобалните разходи за управление и анализ на данни се очаква да нараснат драстично от 185,5 млрд. долара през 2023 г. до 513,3 млрд. долара през 2030 г., с комбиниран годишен темп на растеж от 16%. Не всички компоненти обаче растат с еднакъв темп: анализите растат бързо, докато ръстът на съхранението се забавя. Анализът ще осигури най-бързия растеж в екосистемата за управление на данни: предвижда се той да нарасне от 60,6 млрд. долара през 2023 г. до 227,9 млрд. долара до 2030 г., което представлява съставен годишен темп на растеж от 27%.

С ускоряването на дигитализацията на информацията и бързото нарастване на обема ѝ, управлението на строителни проекти и компании е изправено пред необходимостта от систематично съхраняване, анализиране и обработване на разнообразни, често хетерогенни данни. В отговор на това предизвикателство, от средата на 90-те години на миналия век, индустрията започна масово преминаване към електронно създаване, съхранение и управление на документацията - от електронни таблици и проектни изчисления до чертежи и договори.

Традиционните хартиени документи, които изискват подписи, физическо съхранение, редовна ревизия и архивиране в шкафове, постепенно се заменят с цифрови системи, които съхраняват данни по структуриран начин - в бази данни на специализирани приложения.



## ГЛАВА 1.2.

# ТЕХНОЛОГИИ И СИСТЕМИ ЗА УПРАВЛЕНИЕ В СЪВРЕМЕННОТО СТРОИТЕЛСТВО

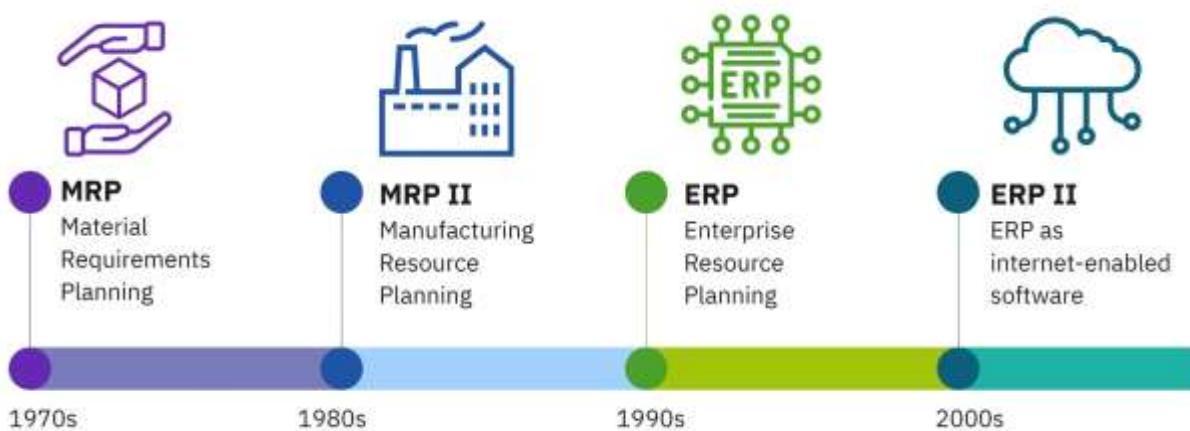
### Цифровата революция и появата на модулни MRP/ERP -системи

Ерата на съвременното цифрово съхранение и обработка на данни започва с появата на магнитната лента през 50-те години на миналия век, която дава възможност за съхранение и използване на големи количества информация. Следващият пробив беше появата на дисковите устройства, които коренно промениха подхода към управлението на данни в строителната индустрия.

С развитието на съхранението на данни голям брой компании навлязоха на пазара на решения и започнаха да разработват модулен софтуер за създаване, съхранение, обработка на данни и автоматизиране на рутинни задачи

Експоненциалното нарастване на информацията и инструментите доведе до нуждата от интегрирани, модулни решения, които не работят с отделни файлове, а помагат за управлението и контрола на потока от данни в различни процеси и проекти.

Първите цялостни инструменти на платформата трябваше не само да съхраняват документи, но и да документират всички заявки за промяна и операции в процесите: кой ги е инициирал, какъв е бил обхватът на заявката и какво е било записано накрая като стойност или атрибут. За тези цели беше необходима система, която да може да проследява точните изчисления и взетите решения (фиг. 1.2-1). Такива платформи бяха първите системи MRP (Material Requirements Planning) и ERP (Enterprise Resource Planning), които станаха популярни от началото на 90-те години на миналия век [14]

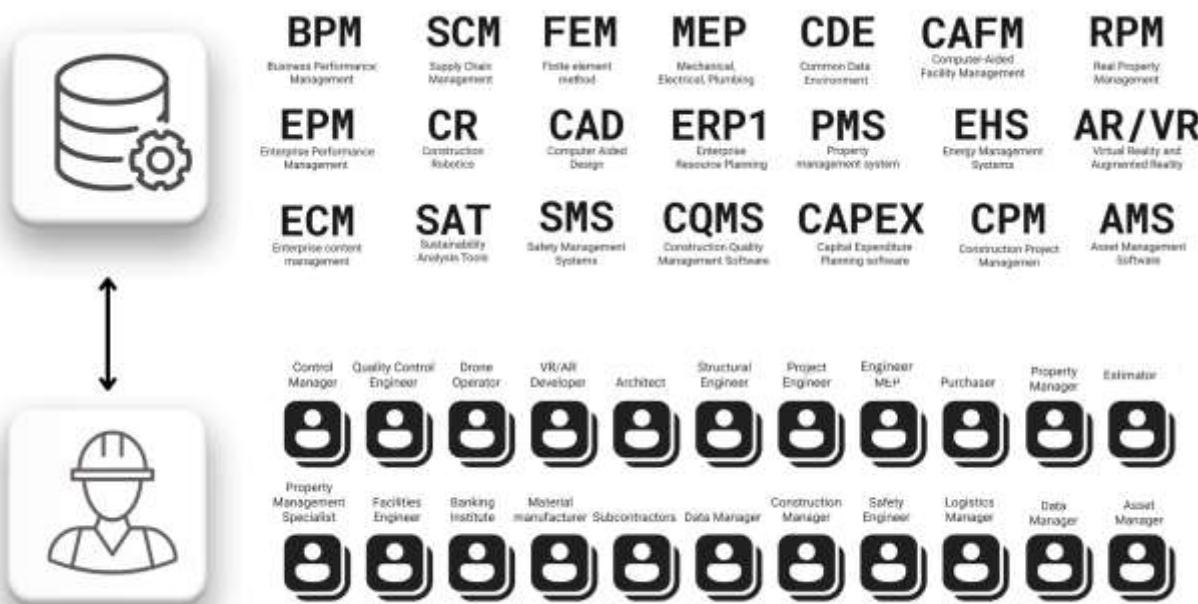


Фигура 1.2-1 Напредъкът в технологиите за съхранение на данни доведе до появата на ERP -системите през 80-те години на миналия век.

Първите системи MRP - и ERP - положиха основите на ерата на цифровизацията в управлението на

бизнес процесите и строителните проекти. Модулните системи, първоначално предназначени за автоматизиране на ключови бизнес процеси, в крайна сметка бяха интегрирани с допълнителни, по-гъвкави и адаптивни софтуерни решения.

Тези допълнителни решения са предназначени за обработка на данни и управление на съдържанието на проекта (фиг. 1.2-2), те или заместват определени модули на големи системи, или ефективно ги допълват, разширявайки функционалността на цялата система.



**Фиг. 1.2-2 Новите софтуерни решения привлякоха в бизнеса армия от мениджъри, които да управляват потоците от данни.**

През последните десетилетия компаниите инвестираха много в модулни системи [15], възприемайки ги като дългосрочни интегрирани решения.

Според доклада на Software Path за 2022 г. [16] средният бюджет на потребител на ERP - система е 9 000 USD. Средно около 26% от служителите на компанията използват такива системи. Така за организация със 100 потребители общите разходи за внедряване на ERP достигат приблизително 900 000 USD.

Инвестициите в патентовани, затворени и модулни решения стават все по-малко оправдани на фона на бързото развитие на модерни, гъвкави и отворени технологии. Ако такива инвестиции вече са направени, важно е обективно да се преоценят ролята на съществуващите системи: дали те остават от съществено значение в дългосрочен план, или функциите им могат да бъдат преразгледани и реализирани по-ефективно и прозрачно.

Един от основните проблеми на съвременните модулни платформи за данни е, че те централизират управлението на данните в рамките на затворени приложения. В резултат на това данните - основният актив на компанията - стават зависими от конкретни софтуерни решения, а не обратното. Това ограничава повторното използване на информацията, усложнява миграцията на и намалява гъвкавостта на бизнеса в бързо променящия се цифров пейзаж.

Ако има вероятност значението или актуалността на затворената модулна архитектура да намалее в бъдеще, има смисъл да се признаят направените днес разходи като невъзвръщаеми и да се съсредоточим върху стратегическото преминаване към по-отворена, мащабируема и адаптивна цифрова екосистема.

Собственият софтуер се характеризира с изключителния контрол на разработчика върху изходния код и потребителските данни, създадени в рамките на използването на такива решения. За разлика от софтуера с отворен код, потребителите нямат достъп до вътрешната структура на приложението и не могат самостоятелно да го преглеждат, променят или адаптират към своите нужди. Вместо това от тях се изисква да закупят лицензи, които предоставят правото да използват софтуера в рамките на ограниченията, определени от доставчика.

Съвременният подход, ориентиран към данните, предлага различна парадигма: данните трябва да се разглеждат като основен стратегически актив - независим, траен и отделен от конкретни софтуерни решения. Приложението, от своя страна, се превръщат в обикновени инструменти за данни, които могат да бъдат заменяни свободно, без рисък от загуба на критична информация.

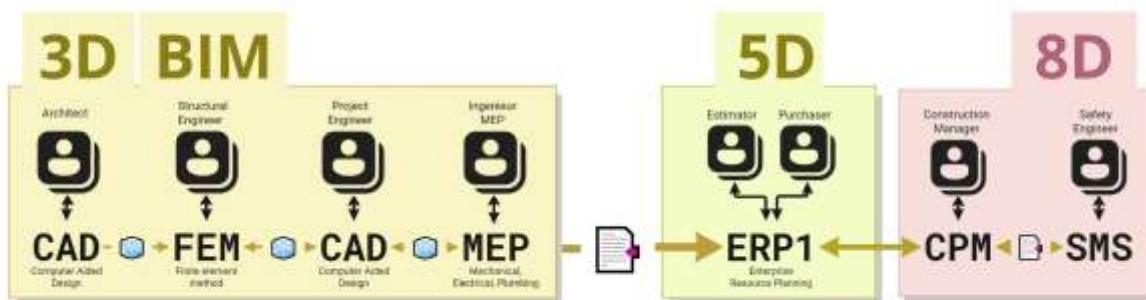
Разработването на ERP и MRP системи през 90-те години на миналия век (Фигура 1.2-1) предостави на предприятията мощни инструменти за управление на процесите, но също така имаше непредвидената последица да увеличи значително броя на хората, които участват в поддържането на информационните потоци. Вместо да автоматизират и опростят оперативните задачи, тези системи често създаваха нови нива на сложност, бюрокрация и зависимост от вътрешните ИТ ресурси.

## **Системи за управление на данни: от извлечане на данни до бизнес предизвикателства**

Съвременните компании са изправени пред необходимостта да интегрират множество системи за управление на данни. Изборът на системи за управление на данни, разумното управление на тези системи и интегрирането на разнородни източници на данни стават от решаващо значение за ефективността на бизнеса.

В средата на 2020-те години можете да откриете стотици (а в големите строителни компании - хиляди) различни системи (Фигура 1.2-3), които трябва да работят в хармония, за да могат всички аспекти на строителния процес да протичат гладко и съгласувано.

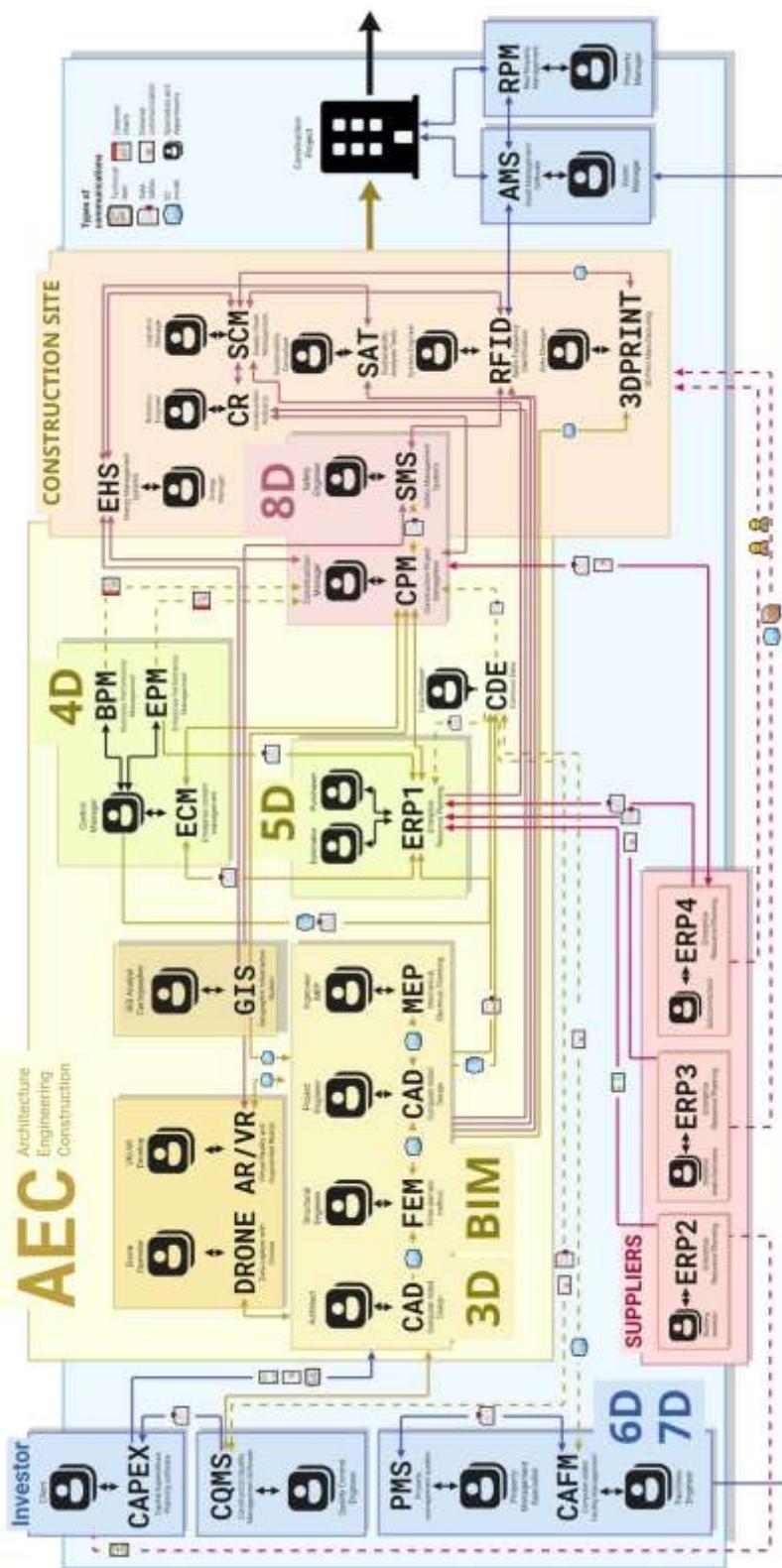
Според проучване на Deloitte от 2016 г.® "Data-Driven Management in Digital Capital Projects" (Управление, базирано на данни, в цифрови капиталови проекти) средностатистическият строителен специалист използва 3,3 софтуерни приложения дневно, но само 1,7 от тях са интегрирани помежду си [17].



Фиг. 1.2-3 Всяка бизнес система изисква професионален екип и отговорен мениджър за качествено управление на данните.

По-долу е представен списък на популярни системи за средни и големи компании в строителния бранш, които се използват за ефективно управление на строителни проекти:

- **ERP (планиране на ресурсите на предприятието)** - осигурява интеграция на бизнес процесите, включително счетоводство, поръчки и управление на проекти.
- **CAPEX (Софтуер за планиране на капиталови разходи)** - използва се за бюджетиране и управление на финансова инвестиции в строителни проекти, помага за определяне на разходите за дълготрайни активи и инвестиции в дългосрочни активи.
- **CAD (Computer-Aided Design) и BIM (Building Information Modeling)** - се използват за създаване на подробни и точни технически чертежи и 3D -модели на проекти. Фокусът на тези системи е върху работата с геометрична информация.
- **MEP (Mechanical, Electrical, Plumbing)** - инженерни системи, които включват механични, електрически и водопроводни компоненти и подробно описват вътрешната "кръвоносна" система на проекта.
- **ГИС (Географски информационни системи)** - използват се за анализ и планиране на терена, включително картография и пространствен анализ.
- **CQMS (софтуер за управление на качеството в строителството)** - гарантира, че строителните процеси са в съответствие с установените стандарти и разпоредби, като помага за отстраняване на дефектите.
- **CPM (управление на строителни проекти)** - включва планиране, координиране и контрол на строителните процеси.
- **CAFM (Computer-Aided Facility Management)** - системи за управление и поддръжка на сгради.
- **Управлението на веригата на доставки (SCM)** е необходимо за оптимизиране на потока от материали и информация между доставчиците и строителната площадка.
- **EPM (Enterprise Performance Management)** - насочен към подобряване на бизнес процесите и ефективността.
- **AMS (Софтуер за управление на активи)** - използва се за оптимизиране на използването, управлението и поддръжката на оборудване и инфраструктура през целия жизнен цикъл на активите.
- **RPM (Управление на недвижимата собственост)** - включва задачи и процеси, свързани с управлението и експлоатацията на сгради и земя, както и на свързаните с тях ресурси и активи.



**Фиг. 1.2-4 Взаимосвързаност на системите, която свързва процесите в компанията с потока от информация между различните отдели.**

- **CAE (Компютърно подпомагано инженерство)** - Компютърно подпомаганото инженерство включва изчислителни и симулационни системи като анализ на крайни елементи (FEA) и изчислителна динамика на флуидите (CFD).
- **CFD (Computational Fluid Dynamics)** - Изчислителна динамика на флуидите, моделиране на флуидни и газови потоци. Подкатегория CAE.
- **CAPP (Компютърно планиране на процеси)** - Компютърно планиране на процеси. Използва се за създаване на карти на маршрути и процеси.
- **CAM (Computer-Aided Manufacturing)** - компютърно подпомогнато производство, създаване на програми за управление на машини с ЦПУ.
- **PDM (Управление на данните за продукта)** - Управление на данните за продукта, система за съхраняване и управление на техническата документация.
- **MES (Manufacturing Execution System)** е система за контрол на производствения процес в реално време.
- **PLM (Product Lifecycle Management)** - управление на жизнения цикъл на елемент от проекта, интегрира PDM, CAPP, CAM и други системи за пълен контрол на продукта от разработвателско до изхвърлянето му.

Тези и много други системи, включващи разнообразни софтуерни решения, са се превърнали в неразделна част от съвременната строителна индустрия (фиг. 1.2-4). В същността си тези системи представляват специализирани бази данни с интуитивни интерфейси, които осигуряват ефективно въвеждане, обработка и анализ на информацията на всички етапи на проектирането и строителството. Интеграцията на цифровите инструменти помежду им не само спомага за оптимизиране на работните процеси, но и значително подобрява точността на решенията, което оказва положително влияние върху сроковете и качеството на изпълнение на проекта.

В половината от случаите обаче няма интеграция. Според статистиката само всяко второ приложение или система е интегрирано с други решения [17]. Това показва продължаващата фрагментация на цифровата среда и подчертава необходимостта от разработване на отворени стандарти и унифицирани интерфейси, за да се осигури обмен на информация от край до край в рамките на строителния проект.

Едно от основните предизвикателства пред интеграцията в съвременните компании остава високата сложност на цифровите системи и изискванията за компетентност на потребителите, необходими за ефективно извлечане и интерпретиране на информация. За поддръжка на всяка внедрена в предприятието система се формира екип от специалисти, начело с ключов мениджър (фиг. 1.2-2).

Ключовият системен мениджър играе решаваща роля за правилното насочване на потока от данни и е отговорен за качеството на крайната информация, точно както първите мениджъри преди хиляди години са били отговорни за числата, написани върху папируси или глинени плочки.

За да се превърнат разнородните информационни потоци в инструмент за управление, от съществено значение е способността за систематично интегриране и управление на данните. В тази архитектура мениджърите трябва да действат като елементи на единна мрежа - като мицел, който свързва отделните части на компанията в цялостен жив организъм, способен да се адаптира и развива.

## Корпоративен мицел: как данните се свързват с бизнес процесите

Процесът на интегриране на данни в приложения и бази данни се основава на обобщаване на информация от различни източници, включително от различни отдели и специалисти (фигура 1.2-4). Специалистите търсят съответните данни, обработват ги и ги прехвърлят в своите системи и приложения за по-нататъшно използване.

Всяка фирмена система, състояща се от набор от инструменти, технологии и бази данни, е дърво на знанието, вкоренено в почвата на историческите данни и растящо, за да даде нови плодове под формата на готови решения: документи, изчисления, таблици, графики и информационни таблица (фиг. 1.2-5). Системите в една компания, подобно на дърветата в определен участък от гората, си взаимодействат и комуникират помежду си, съставлявайки сложна и добре структурирана система, поддържана и управлявана от експертни мениджъри.

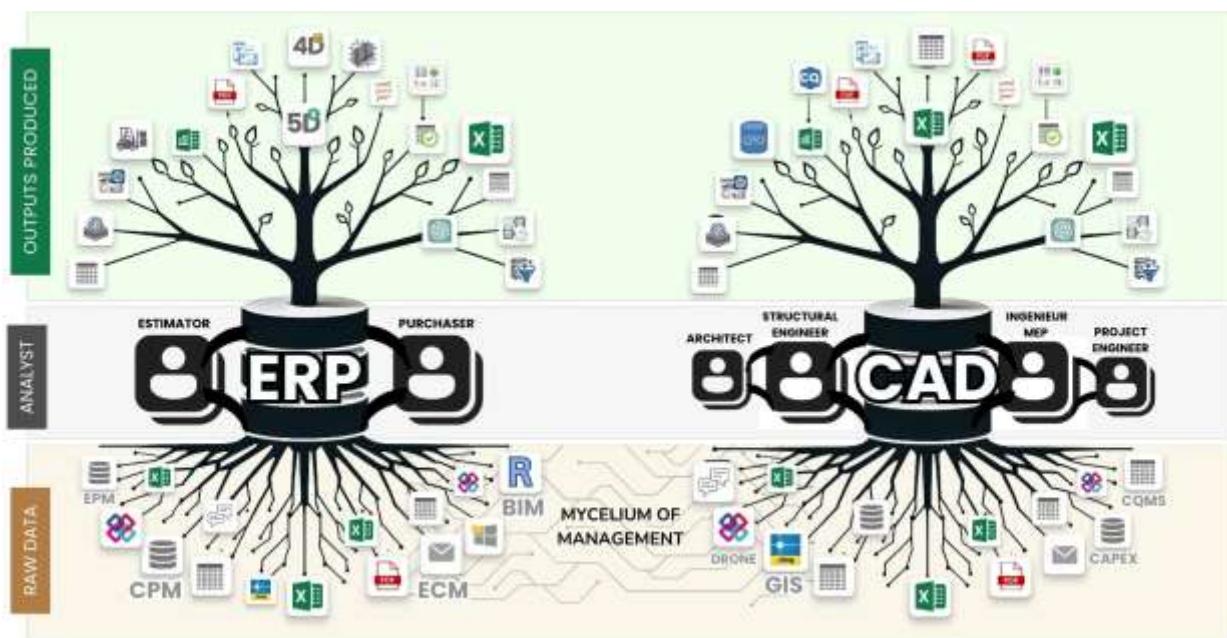
Системата за извлечане и предаване на информация на дадена компания работи като сложна горска мрежа, състояща се от дървета (системи) и мицелни гъби (мениджъри), които действат като проводници и рециклиращи устройства, гарантиращи, че информацията се предава и преминава към правилните системи. Това спомага за поддържането на здравословен и ефективен поток и разпределение на данните в рамките на компанията.

Експертите, подобно на корените, погъщат необработени данни в началните етапи на проекта, като ги превръщат в хранителни вещества за корпоративната екосистема. Системите за управление на данни и съдържание (фиг. 1.2-4 - ERP, CRM, BIM и т.н.) действат като мощни информационни магистрали, по които тези знания циркулират през всички нива на компанията.

Точно както в природата, където всеки елемент от екосистемата играе своята роля, така и в бизнес средата на една компания всеки участник в процеса - от инженера до анализатора - допринася за растежа и плодородието на информационната среда. Тези системни "дървета на данните" (фиг. 1.2-5) не са просто механизми за събиране на информация, а конкурентно предимство, което осигурява устойчивост на компанията.

Горските екосистеми са изненадващо точно отражение на начина, по който са организирани цифровите корпоративни структури. Подобно на стъпаловидната структура на гората - от подрasta до върховете на дърветата - корпоративното управление разпределя задачите по нива на отговорност и функционални отдели.

Дълбоките и разклонени корени на дърветата осигуряват устойчивост и достъп до хранителни вещества. По същия начин солидната организационна структура и стабилните процеси за работа с качествени данни поддържат цялата информационна екосистема на компанията, като допринасят за нейния устойчив растеж и развитие дори в периоди на пазарна нестабилност и кризи (силен вятър).



**Фиг. 1.2-5 Интегрирането на данни чрез различни системи е като мицел, който свързва мениджъри и специалисти в единна информационна мрежа.**

Съвременното разбиране за мащаба в бизнеса се е развило. Днес стойността на една компания се определя не само от нейната видима част - "короните" под формата на заключителни документи и отчети, но и от дълбочината на "кореновата система" от качествено събрани и систематично обработени данни. Колкото повече информация може да бъде събрана и обработена, толкова по-висока става стойността на бизнеса. Компаниите, които методично натрупват "компост" от вече обработени данни и са в състояние да извлекат полезни прозрения от тях, получават стратегическо предимство

Историческата информация се превръща в нов вид капитал, който дава възможност за растеж, оптимизация на процесите и конкурентни предимства. В свят, управляван от данни, печели не този, който има повече, а този, който знае повече.

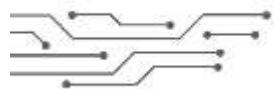
За строителния бранш това означава преминаване към управление на проекти в реално време, при което всички процеси - от проектирането и възлагането на обществени поръчки до координацията на изпълнителите - ще се основават на подходящи, ежедневно актуализирани данни. Интегрирането на информация от различни източници (ERP-системи, CAD-модели, сензори IoT на строителните обекти, RFID) ще позволи да се правят по-точни прогнози, да се реагира бързо на промените и да се избягват забавления, причинени от липсата на актуални данни.

Според доклада Data-Driven Enterprise 2025 на McKinsey & Company (McKinsey & Company®, 2022 [18]) успешните компании на бъдещето ще разчитат на данни във всички ключови аспекти на дейността си - от стратегическите решения до оперативните взаимодействия.

Данните ще престанат да бъдат само инструмент за анализ и ще се превърнат в неразделна част от всички бизнес процеси, осигурявайки прозрачност, контрол и автоматизация на управлението.

Движението, основано на данни под, ще позволи на организациите да сведат до минимум въздействието на човешкия фактор, да намалят оперативните рискове и да увеличат прозрачността и ефективността на вземането на решения.

21-ви век преобръща икономическата парадигма с главата надолу: докато преди петролът се наричаше "черно злато" заради способността му да задвижва машини и транспорт, днес, компресирани под натиска на времето, историческите данни се превръщат в нов стратегически ресурс, който захранва не машини, а алгоритми за вземане на решения, които ще движат бизнеса.



## ГЛАВА 1.3.

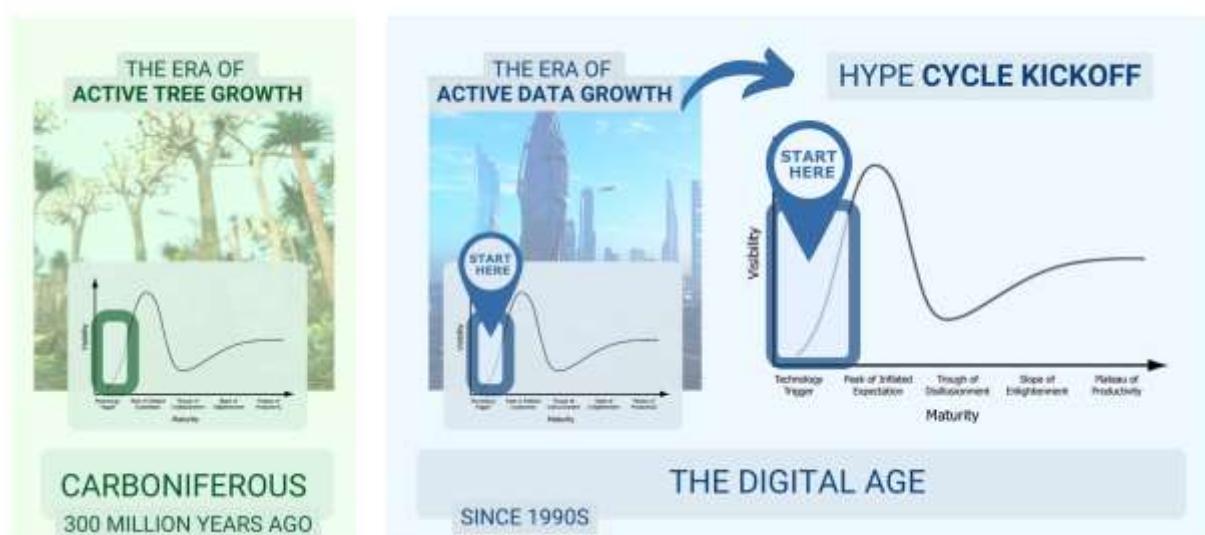
# ЦИФРОВАТА РЕВОЛЮЦИЯ И ЕКСПЛОЗИЯТА НА ДАННИ

### Началото на бума на данните като еволюционна вълна

Строителната индустрия преживява безпрецедентна информационна експлозия. Ако си представим бизнеса като дърво на знанието (фиг. 1.2-5), захранвано с данни, настоящият етап на цифровизация може да се сравни с бързото развитие на растителността през карбонския период - епоха, в която биосферата на Земята се е трансформирала чрез бързото натрупване на биомаса (фиг. 1.3-1).

С глобалната дигитализация обемът на информацията в строителната индустрия се удвоява всяка година. Съвременните технологии позволяват данните да се събират на заден план, да се анализират в реално време и да се използват в мащаби, които само преди малко изглеждаха невъзможни.

Според закона на Мур, формулиран от Гордън Мур (съосновател на Intel®), плътността и сложността на интегралните схеми и количеството обработвани и съхранявани данни се удвояват приблизително на всеки две години [19].



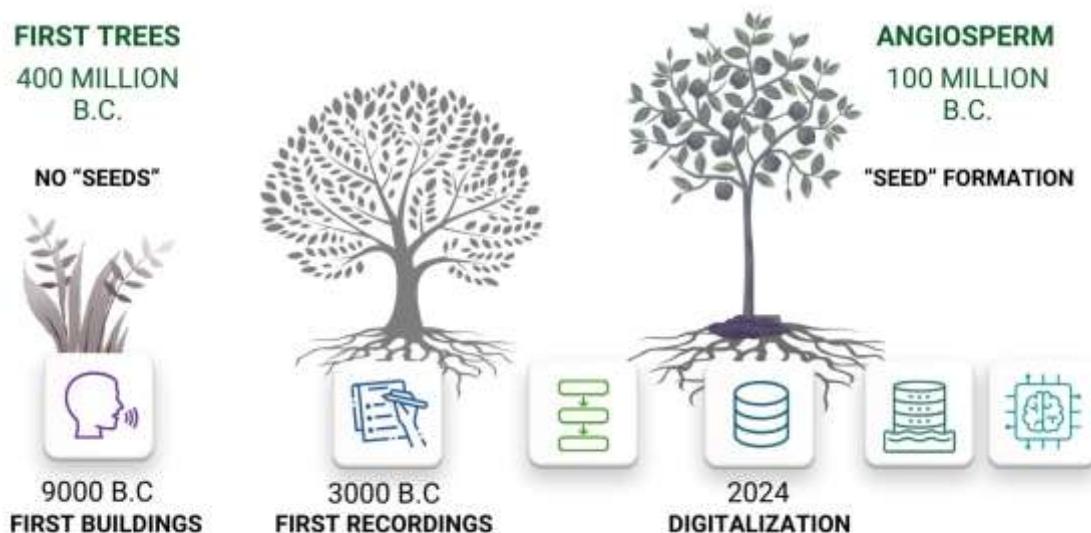
Фиг. 1.3-1 Навлизането на цифровизацията доведе до експоненциално нарастване на данните, подобно на рязкото увеличаване на растителността в епохата на въглищата.

Докато древните мегалитни структури като Гьобекли тепе (Турция) не са оставили документирани знания, подходящи за повторно използване, днес цифровите технологии дават възможност за натрупване и повторно използване на информация. Това може да се сравни с еволюционния преход от спорови растения към семенни растения (Angiosperms): появата на семената е дала началото на широкото разпространение на живота на планетата. (Фиг. 1.3-2).

По подобен начин данните от минали проекти се превръщат в своеобразни "цифрови семена" - носители на ДНК знания, които могат да бъдат увеличавани и използвани в нови проекти и продукти. Появата на съвременни инструменти за изкуствен интелект - машинно обучение и големи

езикови модели (LLM) като ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok - позволява данните да бъдат автоматично извлечани, интерпретирани и прилагани в нови контексти

Точно както семената са предизвикали революция в разпространението на живота на една първоначално безжизнена планета, така и "семената от данни" се превръщат в основа за автоматична появя на нови информационни структури и знания, което позволява на цифровите екосистеми да се развиват самостоятелно и да се адаптират към променящите се изисквания на потребителите.



Фиг. 1.3-2 Цифровите "семена за данни" играят същата еволюционна роля като покритосеменните растения - цъфтящите растения, които са променили екосистемата на Земята.

Стоим на прага на нова ера в строителството, в която експлозията от данни и активното разпространение на "семена от данни" - структурирана информация от минали и текущи проекти - формират основата на цифровото бъдеще на индустрията. Тяхното "опрашване" чрез големи езикови модели на данни (LLM) ни позволява не просто да наблюдаваме цифровите промени, а активно да участваме в създаването на самообучаващи се, адаптивни екосистеми. Това не е еволюция - това е цифрова революция, в която данните се превръщат в основен градивен елемент на една нова реалност

Обемът на данните в строителния бранш се увеличава драстично поради наличието на информация от различни дисциплини през целия жизнен цикъл на строителните проекти. Това огромно натрупване на данни тласка строителната индустрия към ерата на големите данни [20].

- Проф. д-р Ханг Янг, Катедра по строително инженерство и архитектура, Технологичен университет Ухан,

Нарастването на данните в информационната ера напомня на еволюционните процеси в природата: както развитието на горите е променило древния пейзаж на планетата, така и сегашната информационна експлозия променя пейзажа на цялата строителна индустрия.

## Обемът на данните, генериирани в една съвременна компания

През последните две години са създадени 90% от всички съществуващи данни в света [21]. От 2023 г. всеки човек, включително специалистите в строителната индустрия, генерира около 1,7 мегабайта данни в секунда [22], а общото количество данни в света ще достигне 64 зетабайта през 2023 г. и се очаква да надхвърли 180 зетабайта, или  $180 \cdot 10^{15}$  мегабайта, до 2025 г. [23].

Тази информационна експлозия има исторически прецедент - изобретяването на печатарската преса от Йоханес Гутенберг през XV век. Само петдесет години след въвеждането ѝ броят на книгите в Европа се удвоява: за няколко десетилетия са отпечатани толкова книги, колкото са били създадени на ръка през предходните 1200 години [24]. Днес сме свидетели на още по-бърз растеж: обемът на данните в света се удвоява на всеки три години.

Предвид сегашния темп на нарастване на данните, строителната индустрия има потенциала да генерира толкова информация през следващите няколко десетилетия, колкото е натрупала през цялата си предишна история

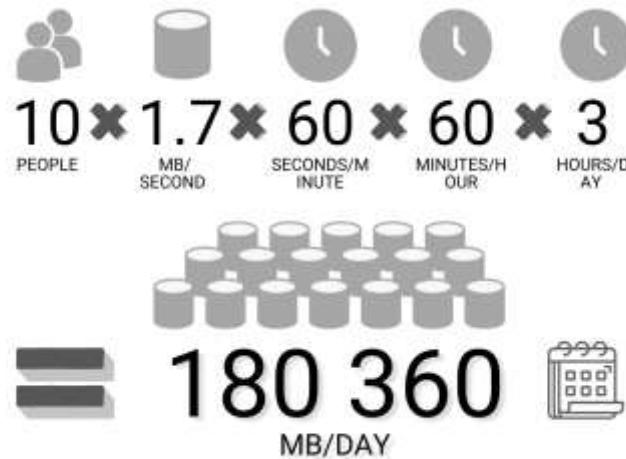


Фиг. 1.3-3 Ежедневното съхраняване на данни от всеки служител на сървърите на компанията допринася за постоянно нарастване на обема на данните.

В днешния свят на строителния бизнес дори малките компании генерират ежедневно огромно количество многоформатна информация, а цифровият отпечатък дори на малка строителна компания може да достигне десетки гигабайти на ден - от модели и чертежи до фотографски записи и

сензори на обекта. Ако приемем, че всеки техник генерира средно около 1,7 MB данни в секунда, това е равносилно на около 146 GB на ден или 53 TB на година (фиг. 1.3-3).

Когато екип от 10 души работи активно само 3 часа дневно, кумулативното количество информация, генерирано на ден, достига 180 гигабайта (фиг. 1.3-4).



**Фигура 1.3-4 Компания с 10 души генерира приблизително 50-200 гигабайта данни на ден.**

Ако приемем, че 30% от работните данни са нови (останалите се презаписват или изтриват), фирма с 10 души може да създава по няколкостотин гигабайта нови данни на месец (действителните цифри зависят от вида на дейността на компанията)

Следователно е ясно, че не просто генерираме все повече данни, а се сблъскваме с нарастваща нужда от тяхното ефективно управление, съхранение и дългосрочна наличност. И докато преди данните можеха да "лежат" на локални сървъри без никакви разходи, в контекста на цифровата трансформация все повече компании започват да използват облачни решения като основа на информационната си инфраструктура.

### **Разходите за съхранение на данни: икономически аспект**

През последните години все повече компании възлагат съхранението на данни на облачни услуги. Например, ако една компания хоства половината от данните си в облака, при средна цена от 0,015 USD за гигабайт на месец, разходите ѝ за съхранение могат да се увеличат с 10-50 USD [25] всеки месец.

За малка компания с типични модели на генериране на данни разходите за съхранение в облак могат да варират от стотици до потенциално над хиляда долара на месец (Фигура 1.3-5) в няколко

години, което създава потенциално значителна финансова тежест.

Според проучването на Forrester "Предприятията възлагат съхранението на данни на външни изпълнители, тъй като сложността нараства" [26], в което са анкетирани 214 ръководители на технологични инфраструктури. [26], в което са анкетирани 214 ръководители на технологични инфраструктури, повече от една трета от организациите възлагат съхранението на данни на външни изпълнители, за да се справят с нарастващия обем и сложност на операциите с данни, като почти две трети от предприятията предпочитат абонаментен модел.



Фиг. 1.3-5 Преместването на данни в облака може да увеличи месечните разходи за съхранение с до 2000 долара, дори за компания с едва 10 служители.

Ситуацията се усложнява допълнително от ускореното внедряване на технологии, базирани на облачни технологии, като CAD (BIM), CAFM, PMIS и ERP -системи, които допълнително увеличават разходите за съхранение и обработка на данни. В резултат на това компаниите са принудени да търсят начини за оптимизиране на разходите и намаляване на зависимостта от доставчиците на облачни услуги.

От 2023 г. насам, с активното разработване на големи езикови модели (LLM), подходите към съхранението на данни започнаха да се променят. Все повече компании се замислят да си върнат контрола върху своите данни, тъй като става по-безопасно и по-изгодно да обработват информацията на собствените си сървъри.

В този контекст на преден план излиза тенденцията за отказ от облачно съхранение и обработка само на необходимите данни в полза на локално внедряване на корпоративни LLM и AI -решения. Както посочва главният изпълнителен директор на Microsoft в едно от своите интервюта [27], вместо да се разчита на няколко отделни приложения или на облачни SaaS за изпълнение на различни задачи, агентите на ИИ ще управяват процесите в базите данни, автоматизирайки функциите на различни системи.

[...] Старият подход към този проблем [обработка на данни] беше следният: ако си спомним как различните бизнес приложения се справяха с интеграцията, те използваха конектори. Компаниите продаваха лицензи за тези конектори и бизнес моделът се формираше около това. SAP [ERP] е един от класическите примери: достъпът до данните на SAP е възможен само ако имате правилния конектор. Така че ми се струва, че нещо подобно ще се появи и в случая с взаимодействието между агентите [AI] [...]. Подходът, поне този, който ние възприемаме, е следният: мисля, че концепцията за съществуването на бизнес приложения вероятно ще се срине в ерата на агентите с [ИИ]. Защото, ако се замислите, те по същество са бази данни с куп бизнес логика

- Сатя Надела, главен изпълнителен директор на Microsoft, интервю за канал BG2, 2024 г. [28]

В тази парадигма подходът LLM, основан на данни, надхвърля класическите системи. Изкуственият интелект се превръща в посредник между потребителя и данните (фиг. 2.2-3, фиг. 2.2-4), като елиминира необходимостта от множество междинни интерфейси и повишава ефективността на бизнес процесите. Ще говорим повече за този подход към работата с данни в главата "Превръщане на хаоса в ред и намаляване на сложността".

Докато архитектурата на бъдещето все още се оформя, компаниите вече се сблъскват с последствията от минали решения. Машабната цифровизация през последните десетилетия, придружена от въвеждането на разнородни системи и неконтролирано натрупване на данни, доведе до нов проблем - информационно претоварване.

## **Граници на натрупването на данни: от маса към значение**

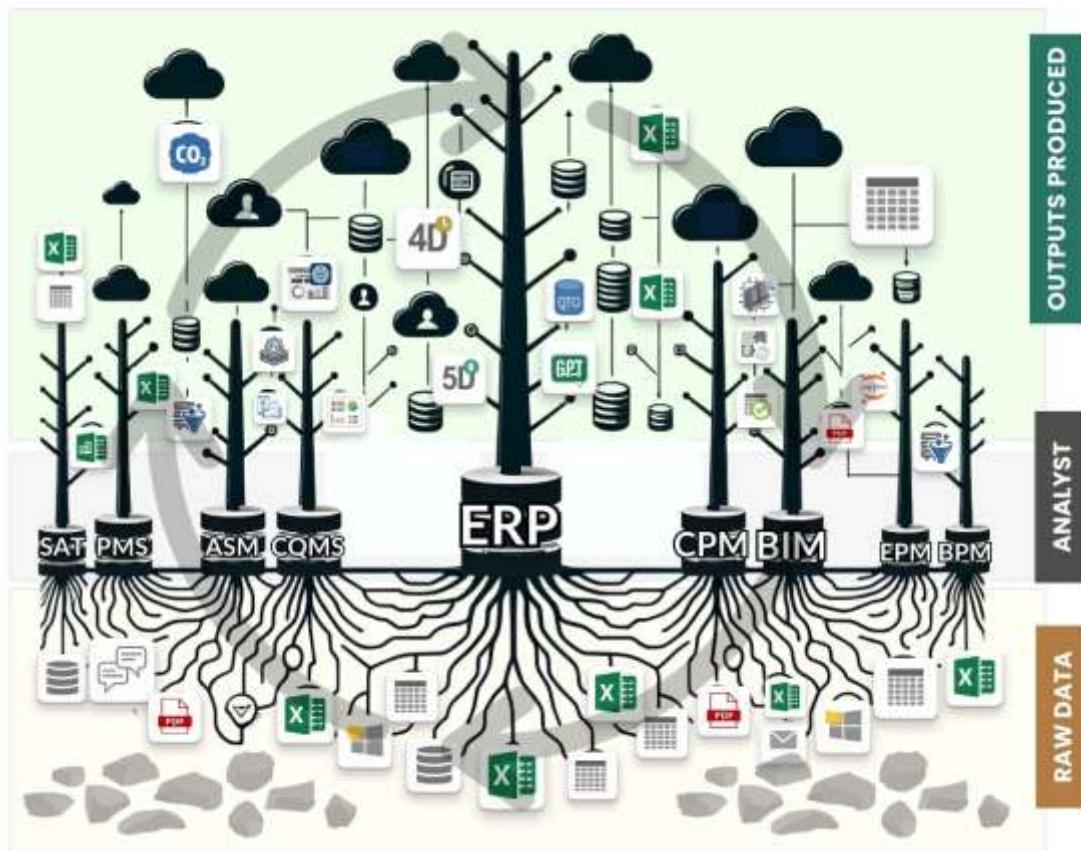
Съвременните фирмени системи се развиват и функционират успешно в условията на управляем растеж, когато обемът на данните и броят на приложенията са в баланс с възможностите на ИТ отделите и мениджърите. През последните десетилетия обаче цифровизацията доведе до неконтролирано нарастване на обема и сложността на данните, което предизвика ефект на пренасищане в информационната екосистема на компаниите.

Днес сървърите и съръдженятията за съхранение са подложени на безprecedентен наплив от необработена и многоформатна информация, която няма време да се превърне в компост и бързо губи значение. Ограничните ресурси на компанията не могат да се справят с този поток и данните се натрупват в изолирани силози (т.нар. "силози"), които изискват ръчна обработка, за да се извлече полезна информация.

В резултат на това, подобно на гора, обрасла с бръшлян и покрита с мухъл, съвременният системи за управление на компаниите често страдат от информационно претоварване. Вместо да подхранва информационния хумус в сърцето на корпоративната екосистема, се образуват изолирани области от данни с различен формат, което неизбежно води до намаляване на общата ефективност на бизнес процесите.

Дългият период на експоненциално нарастване на данните през последните 40 години неминуемо ще бъде последван от фаза на насищане и последващо охлажддане. Когато съхранението на данни достигне своя предел, ще настъпи качествена промяна: данните вече няма да бъдат просто обект за съхранение, а стратегически ресурс.

С развитието на изкуствения интелект и машинното обучение компаниите имат възможност да намалят разходите за обработка на информация и да преминат от количествен растеж към качествено използване на данните. През следващото десетилетие строителният бранш ще трябва да измести фокуса си от създаването на все повече данни към осигуряването на тяхната структура, цялост и аналитична стойност.



Фигура 1.3-6 Изолираните източници на данни възпрепятстват обмена на информация между системите за данни.

Основната стойност вече не е в количеството информация, а в способността за автоматичното ѝ интерпретиране и превръщането ѝ в приложни знания, полезни за вземането на управлениски решения. За да станат данните наистина полезни, те трябва да се управляват правилно: да се събират, проверяват, структурират, съхраняват и анализират в контекста на конкретни бизнес задачи.

Процесът на анализ на данни в една компания прилича на цикъла на живот и разпад на дърветата в гората и появата на нови млади и силни дървета: зрелите дървета умират, разлагат се и стават хранителна среда за нов растеж. Завършените и приключили процеси, когато бъдат завършени, стават част от информационната екосистема на компанията и в крайна сметка се превръщат в информационен хумус, който подхранва бъдещия растеж на нови системи и данни.

На практика обаче този цикъл често се нарушава. Вместо органично обновяване се формира слоест хаос, подобен на геологки пластове, при който нови системи се наслагват върху стари без дълбока интеграция и структуриране. В резултат на това се появяват разнородни информационни "силози", които възпрепятстват циркуляцията на знанията и усложняват управлението на данните.

## **Следващи стъпки: от теория на данните към практическа промяна**

Еволюцията на данните в строителството е път от глинени плочки до съвременни модулни платформи. Предизвикателството днес не е да се събира информация, а да се създаде рамка, която превръща разнородните и разнообразни данни в стратегически ресурс. Независимо дали ролята ви е на мениджър на компания или на инженер, разбирането на стойността на данните и как да работите с тях ще бъде ключово умение в бъдеще.

За да обобщим тази част, си струва да подчертаем основните практически стъпки, които ще ви помогнат да прилагате обсъжданите подходи в ежедневните си задачи:

- Извършване на личен одит на информационните потоци
  - Направете списък на всички системи и приложения, с които работите ежедневно.
  - Отбележете местата, където прекарвате най-много време в търсене или повторна проверка на данни.
  - Определете основните си източници на информация
  - Анализирайте текущия си пейзаж от приложения за излишък и дублиране на функции
- Стремеж към преминаване през процесите по нива на аналитична зрялост
  - Започнете задачите си с описателен анализ (какво се случи?)
  - Постепенно въведете диагностика (защо се е случило това?)
  - Помислете как в процесите можете да преминете към прогнозен (какво ще се случи?) и предписващ (какво да се направи?) анализ.
- Започнете да структурирате работните си данни
  - Внедряване на единна система за именуване на файлове и папки, които често

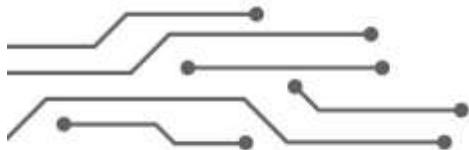
използвате в работата си

- Създаване на шаблони за често използвани документи и отчети
- Редовно архивиране на завършени проекти с ясна структура

Дори и да не можете да промените цялата информационна инфраструктура в екипа или компаниета си, започнете със собствените си процеси и малки подобрения в ежедневната си работа. Не забравяйте, че истинската стойност на данните не е в техния обем, а в способността да се извлекат от тях полезни идеи. Дори малки, но правилно структурирани и анализирани набори от данни могат да окажат значително въздействие, когато бъдат интегрирани в процесите на вземане на решения.

В следващите части на книгата ще преминем към конкретни методи и инструменти за работа с данни, ще разгледаме начините за трансформиране на неструктурата информация в структурирани масиви, ще разгледаме технологите за автоматизация на анализите и ще опишем подробно как да изградим ефективна аналитична екосистема в строителна компания.





## II ЧАСТ

# КАК СТРОИТЕЛНИЯТ БИЗНЕС ПОТЪВА В ХАОС ОТ ДАННИ

Втората част е посветена на критичен анализ на предизвикателствата, пред които са изправени строителните компании при обработката на нарастващи обеми от данни. Подробно са разгледани последиците от разпокъсаността на информацията и явлението "данни в силози", което възпрепятства ефективното вземане на решения. Изследвани са проблемите на HiPPO -подхода (Highest Paid Person's Opinion) и неговото въздействие върху качеството на управленските решения в строителните проекти. Оценява се влиянието на динамичните бизнес процеси и тяхната нарастваща сложност върху информационните потоци и оперативната ефективност. Дадени са конкретни примери за това как прекомерната сложност на системите увеличава разходите и намалява гъвкавостта на организацията. Специално внимание се обръща на ограниченията, създадени от патентованите формати, и на перспективите за използване на отворени стандарти в строителната индустрия. Представена е концепцията за преминаване към софтуерни екосистеми, базирани на изкуствен интелект и LLM, които свеждат до минимум прекомерната сложност и техническите бариери.

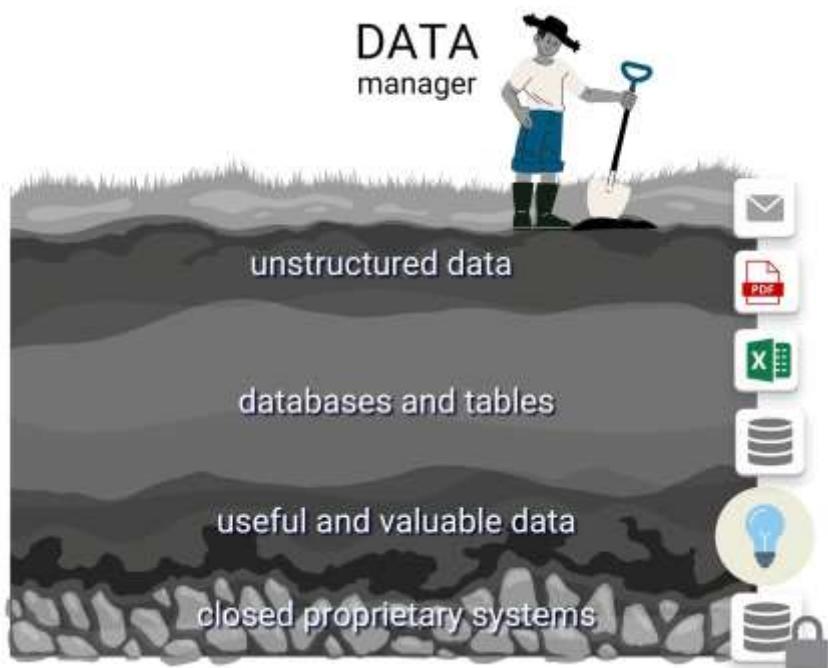
## ГЛАВА 2.1.

### ФРАГМЕНТАЦИЯ НА ДАННИТЕ И "СИЛОЗИ"

#### Колкото повече инструменти, толкова по-ефективен е бизнесът?

На пръв поглед може да изглежда, че повече цифрови инструменти водят до по-голяма ефективност. На практика обаче това не е така. С всяко ново решение, независимо дали става дума за облачна услуга, наследена система или поредния отчет в Excel, компанията добавя още един слой към своя цифров пейзаж - слой, който често не е интегриран с останалите (фигура 2.1-1).

Данните могат да се оприличат на въглища или нефт: натрупването им отнема години, уплътнени под пластове хаос, грешки, неструктурирани процеси и забравени формати. За да извлекат наистина полезна информация от тях, компаниите трябва буквально да се проврат през пластове от остарели решения и дигитален шум.



**Фиг. 2.1-1** Разнообразните данни образуват разченени слоеве - дори "златните" прозрения се губят в геологическите скали на системната сложност.

Всяко ново приложение оставя след себе си следа: файл, таблица или цял изолиран "силоз" на сървъра. Единият слой е глина (остарели и забравени данни), другият е пясък (разнородни таблици и отчети), а третият е гранит (затворени патентовани формати, които не могат да бъдат интегрирани). С течение на времето цифровата среда на компанията все повече прилича на резервоар за неконтролирано натрупване на информация, където стойността се губи дълбоко в сървърите на компанията.

С всеки нов проект и всяка нова система се усложнява не само инфраструктурата, но и пътят до

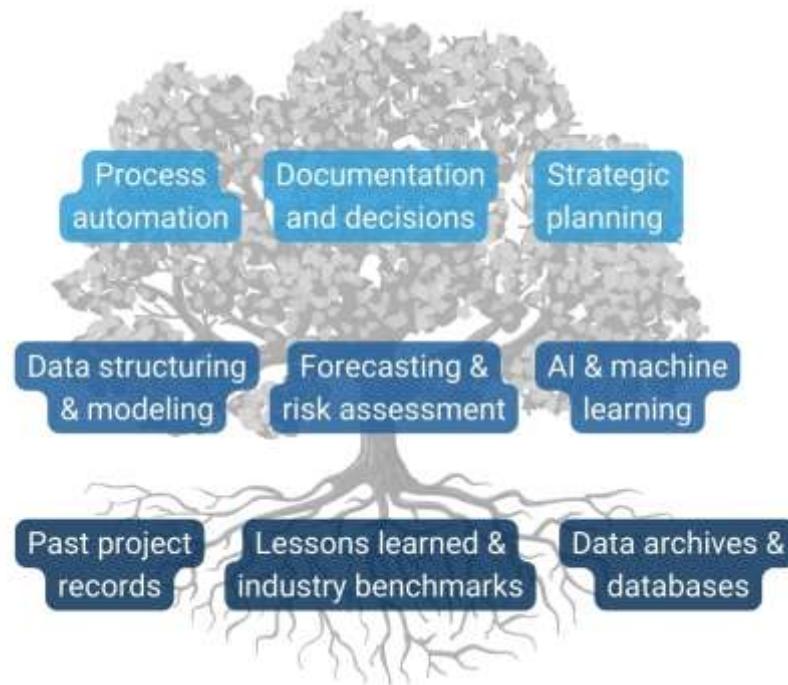
качествени и полезни данни. Достигането до ценната "скала" изисква дълбоко прочистване, структуриране на информацията, "раздробяване", групиране в смислени парчета и извличане на стратегически важни прозрения чрез анализ и моделиране на данни.

*Данните са ценно нещо и те ще продължат да съществуват по-дълго от самите системи [които обработват данните] [29].*

- Тим Бърнърс-Лий, бащата на световната мрежа и създател на първия уебсайт

Преди данните да се превърнат в "ценност" и надеждна основа за вземане на решения, те трябва да бъдат внимателно подгответи. Именно подходящата предварителна обработка превръща разнородните данни в структуриран опит, полезен информационен хумус, който след това се превръща в инструмент за прогнозиране и оптимизация.

Съществува погрешно схващане, че за да започнете да анализирате, се нуждаете от идеално чисти данни, но на практика умението да работите с мръсни данни е съществена част от процеса.



**Фиг. 2.1-2** Данните са кореновата система и основата на бизнеса, който от своя страна се основава на процесите на вземане на решения.

Тъй като технологиите продължават да се развиват, вашият бизнес също трябва да върви напред и да се научи как да създава стойност от данните. Точно както компаниите за добив на нефт и

въглища изграждат инфраструктурата за извлечане на полезни изкопаеми, така и предприятията трябва да се научат как да управляват потока от нова информация на собствените си сървъри и да извлечат ценна информация от неизползвани, неформатирани и остарели данни, превръщайки ги в стратегически ресурс.

Създаването на полета (складове за данни) е първата стъпка. Дори и най-мощните инструменти не решават проблема с изолацията на данните и многоформатните данни, ако компаниите продължават да работят в изолирани системи. Когато данните съществуват отделно едни от други, без да се пресичат и да се обменя информация, предприятията са изправени пред ефекта на "силоzите за данни". Вместо единна, последователна инфраструктура, компаниите са принудени да изразходват ресурси за обединяване и синхронизиране на данни.

## **Силоzите от данни и тяхното въздействие върху ефективността на компанията**

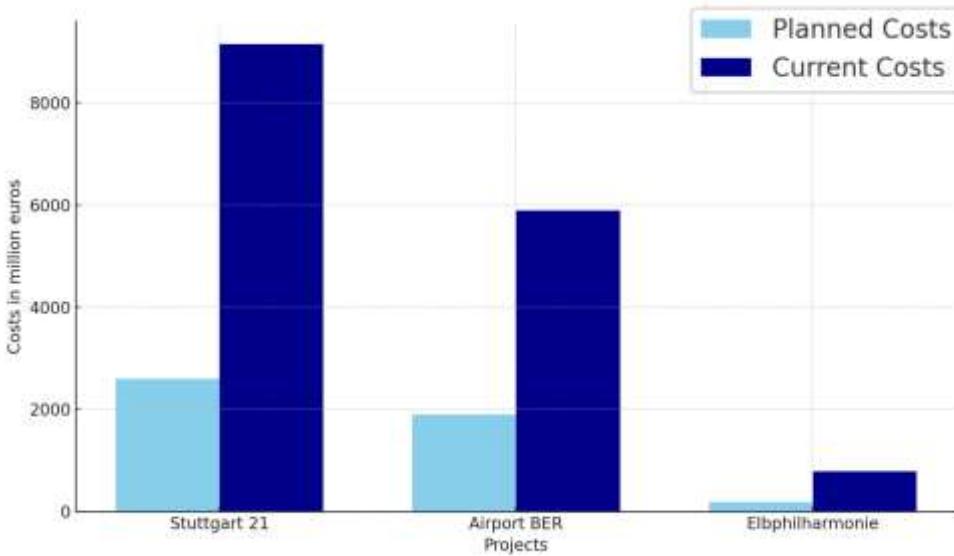
Представете си, че строите жилищен комплекс, но всеки екип има свой собствен проект. Едни строят стени, други полагат комуникации, а трети прокарват пътища, без да се консултират помежду си. В резултат на това тръбите не съвпадат с отворите в стените, асансьорните шахти не съответстват на етажите, а пътищата трябва да се демонтират и полагат наново.

Тази ситуация не е само хипотетичен сценарий, а реалност за много съвременни строителни проекти. Поради големия брой генерални и подизпълнители, работещи с различни системи и без единен координационен център, процесът се превръща в поредица от безкрайни одобрения, преработки и конфликти. Всичко това води до значителни закъснения и многократно увеличаване на разходите по проекта.

Класическата ситуация на строителната площадка е проста: кофражът е готов, но доставката на арматура не е пристигнала навреме. При проверка на информацията в различните системи комуникацията е приблизително следната:

- ❶ **Майсторът на строителната площадка** на 20-ия ден пише на ръководителя на проекта: "Приключи с поставянето на кофража, къде е арматурата?"
- ❷ **Ръководител на проекта (PMIS)** към отдела за обществени поръчки: - "Кофражът е готов. В моята система [PMIS] пише, че арматурното желязо е трябвало да пристигне на 18-ти. Къде са арматурните заготовки?"
- ❸ **Специалист по веригата на доставки (ERP)**: - "Според нашата ERP система доставката ще бъде на 25-ти".
- ❹ **Инженер по данните** или ИТ отдел (отговарящ за интеграциите): - В PMIS датата е 18-то число, а в ERP - 25-то число. Няма връзка с OrderID между ERP и PMIS, така че данните не са синхронизирани. Това е типичен пример за информационен пропуск.
- ❺ **Ръководител на проекта** към главен **менеджър** - "Доставката на фитинги се забави, обектът стои и не е ясно кой носи отговорност".

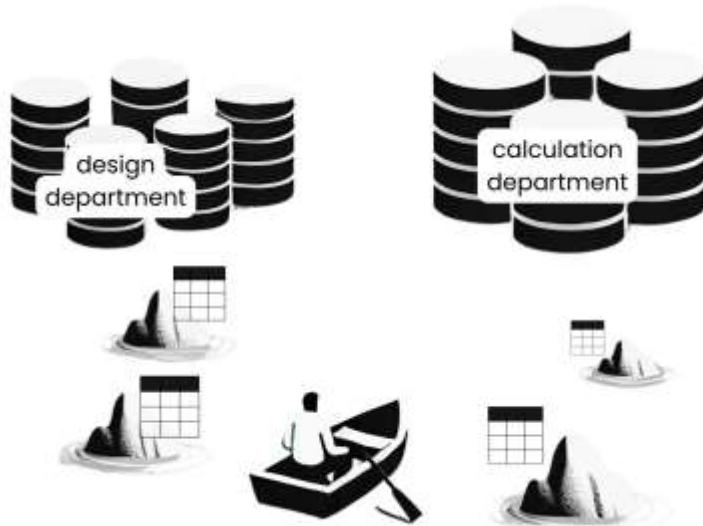
Причината за инцидента е изолирането на данни в различни системи. Чрез интегриране и обединяване на източниците на данни, създаване на единно хранилище на информация и автоматизиране чрез ETL-инструменти (Apache NiFi, Airflow или n8n), силоzите между системите могат да бъдат премахнати. Тези и други методи и инструменти ще бъдат разгледани подробно в следващите раздели на книгата.



**Фиг. 2.1-3 Сравнение на планираните и действителните разходи за големи инфраструктурни проекти в Германия.**

Същото се случва и с корпоративните системи: първо се създават изолирани решения, а след това се налага да се изразходват огромни бюджети за тяхното интегриране и хармонизиране. Ако моделите за данни и комуникация бяха обмислени от самото начало, нямаше да има нужда от интеграция. Отделните данни създават хаос в цифровия свят, подобно на некоординиран строителен процес.

Според проучването на KPMG за 2023 г. "Строителство 4.0: време за промяна" само 36% от компаниите споделят ефективно данни между отделите, а 61% са изправени пред сериозни проблеми поради изолирани "силози" за данни [30].



**Фиг. 2.1-4 Години наред трудни за събиране данни се натрупват в изолирани "силози" за съхранение с риск никога да не ще са използвани.**

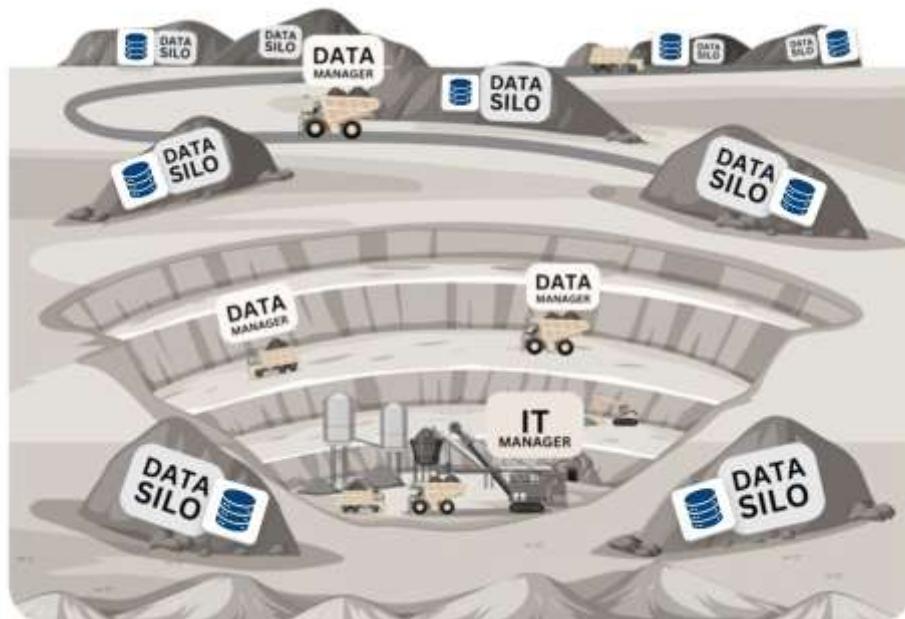
Данните на компанията се съхраняват в изолирани системи, подобно на отделни дървета, разпръснати из пейзажа. Всяко от тях съдържа ценна информация, но липсата на връзки между тях не позволява създаването на единна, взаимосвързана екосистема. Тази изолираност пречи на потока от данни и ограничава способността на организацията да вижда пълната картина. Свързването на тези силози е изключително дълъг и сложен процес на отглеждане на мицел на гъби на управлениеско ниво, за да се научите как да прехвърляте отделни части от информацията между системите.

Според проучване на WEF от 2016 г. една от основните пречки пред цифровата трансформация е липсата на общи стандарти за данни и разположеността.

*Строителната индустрия е една от най-фрагментираните в света и зависи от гладкото взаимодействие на всички участници във веригата на стойността [5].*

- Световен икономически форум 2016: Формиране на бъдещето на

Дизайнерите, мениджърите, координаторите и разработчиците често предпочитат да работят самостоятелно, за да избегнат сложността на координацията. Тази естествена склонност води до създаването на информационни "силози", в които данните са изолирани в рамките на отделни системи. Колкото повече такива изолирани системи има, толкова по-трудно е да се накарат да работят заедно. С течение на времето всяка система се сдобива със собствена база данни и специализиран отдел за поддръжка от мениджъри (фигура 1.2-4), което допълнително усложнява интеграцията.



**Фиг. 2.1-5** Всяка система има склонност да създава свои собствени уникални силози от данни, които трябва да се обработват с подходящи инструменти [31].

Омагьосаният кръг в корпоративните системи изглежда така: компаниите инвестират в сложни ISO

леатирани решения, след което се сблъскват с високи разходи за тяхната интеграция, а разработчиците, оствъзвайки сложността на комбинирането на системите, предпочитат да работят в своите затворени екосистеми. Всичко това увеличава фрагментацията на ИТ ландшафта и затруднява миграцията към нови решения (фигура 2.1-5). В крайна сметка мениджърите критикуват силозите от данни, но рядко анализират причините за тях и начините за предотвратяването им. Мениджърите се оплакват от остарелите ИТ системи, но тяхната замяна изисква значителни инвестиции и рядко дава очакваните резултати. В резултат на това дори опитите за борба с проблема често влошават положението.

Основната причина за разрива е отдаването на приоритет на приложенията пред данните. Компаниите първо разработват отделни системи или купуват готови решения от доставчици, а след това се опитват да ги обединят, като създават дублиращи се и несъвместими хранилища и бази данни.

Преодоляването на проблема с фрагментацията изисква радикално нов подход - приоритет на данните пред приложенията. Компаниите трябва първо да разработят стратегии за управление на данните и модели на данните, а след това да изградят системи или да закупят решения, които да работят с единен набор от информация, а не да създават нови бариери.

*Навлизаме в нов свят, в който данните може би са по-важни от софтуера.*

- Тим О'Райли, главен изпълнителен директор на O'Reilly Media, Inc.

Проучването на McKinsey Global Institute "Преосмисляне на строителството: пътят към подобрена производителност" (2016 г.) показва, че строителната индустрия изостава от други сектори в цифровата трансформация [32]. Според доклада внедряването на автоматизирано управление на данни и цифрови платформи може значително да подобри производителността и да намали загубите, свързани с непоследователност на процесите. Необходимостта от цифрова трансформация се подчертава и в доклада на Egan (Великобритания, 1998 г.) [33], в който се изтъква ключовата роля на интегрираните процеси и съвместния подход в строителството.

В резултат на това, докато през последните 10 000 години основният проблем за мениджърите на данни е бил липсата на данни, с лавинообразното нарастване на данните и системите за управление на данни потребителите и мениджърите се сблъскват с проблем - свръхнасищане на данни, което затруднява намирането на законосъобразна и качествена информация.

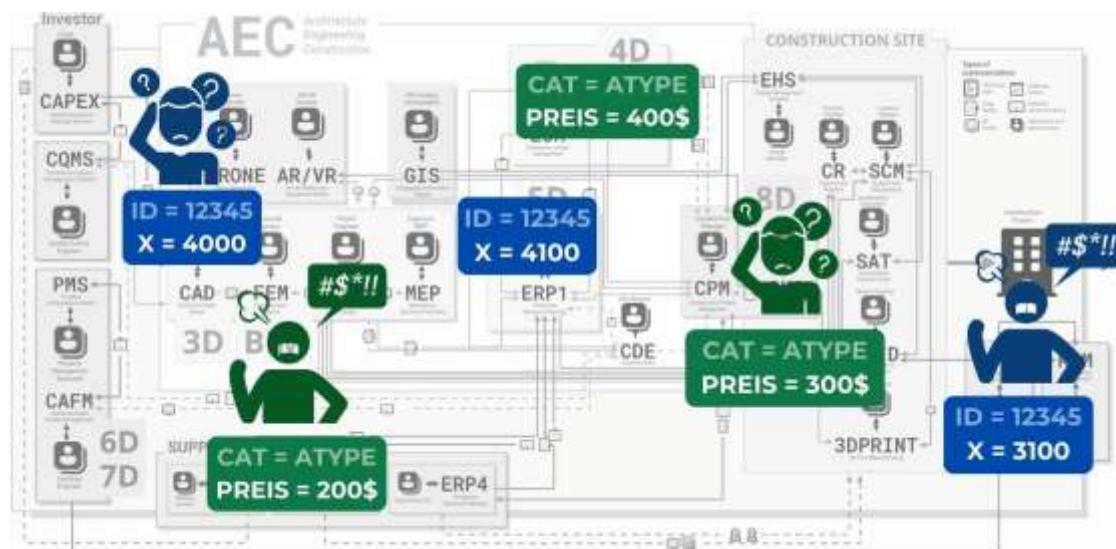
Разнородните силози от данни неизбежно водят до сериозния проблем с пониженото качество на данните. При множество независими системи едни и същи данни могат да съществуват в различни версии, често с противоречиви стойности, което създава допълнителни трудности за потребителите, които трябва да определят коя информация е подходяща и надеждна.

## Дублиране и липса на качество на данните като последица от разединението

Поради проблема със силозите от данни мениджърите трябва да отделят значително време за търсене и съгласуване на данни. За да се предпазят от проблеми с качеството, компаниите създават сложни структури за управление на информацията, в които вертикал от мениджъри отговаря за търсенето, проверката и съгласуването на данните. Този подход обаче само увеличава бюрокрацията и забавя вземането на решения. Колкото повече са данните, толкова по-трудно е те да бъдат анализирани и интерпретирани, особено ако няма единен стандарт за тяхното съхранение и обработка.

С множеството софтуерни приложения и системи, които растат като гъби след дъжд през последното десетилетие, проблемът със силозите и неподходящото качество на данните става все по-важен за крайните потребители. Едни и същи данни, но с различни стойности, вече могат да бъдат открити в различни системи и приложения (фигура 2.1-6). Това води до трудности за крайните потребители, когато се опитват да определят коя версия на данните е подходяща и правилна сред многото налични. Това води до грешки при анализите и в крайна сметка при вземането на решения.

За да се предпазят от проблеми с намирането на правилните данни, мениджърите на компанията създават многостепенна бюрокрация от мениджъри по проверката. Тяхната задача е да могат бързо да намират, проверяват и изпращат необходимите данни под формата на таблици и отчети, като се ориентират в лабиринта от различни системи.

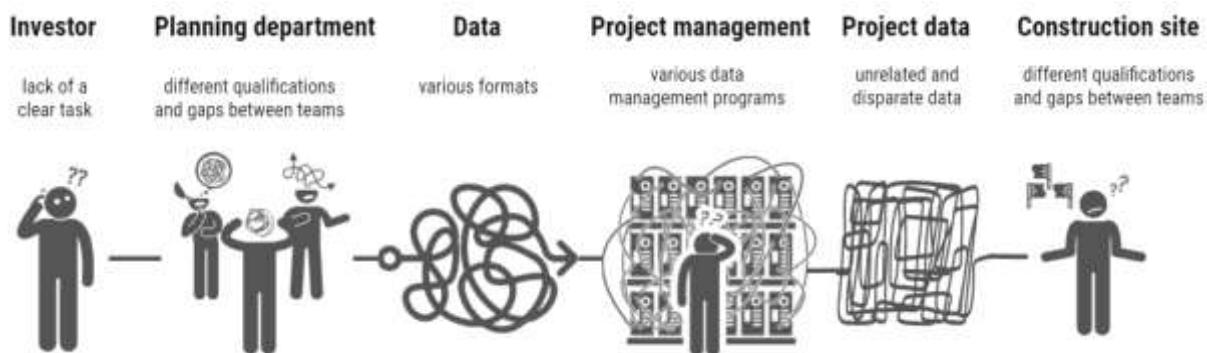


Фиг. 2.1-6 В опитите си да намерят правилните данни, мениджърите трябва да гарантират качеството и правната надеждност на данните между различните системи.

На практика обаче този модел поражда нови трудности. Когато данните се управляват ръчно и информацията е разпръсната в много несвързани решения, всеки опит за получаване на точна и актуална информация чрез пирамидата от лица, вземащи решения (фигура 2.1-7), се превръща в тясно място - отнемащо време и склонно към грешки.

Ситуацията се влошава от лавинообразното навлизане на цифрови решения. Пазарът на софтуер продължава да бъде наводняван от нови инструменти, които изглеждат обещаващи. Но без ясна

стратегия за управление на данните, тези решения не се интегрират в единна система, а вместо това създават допълнителни слоеве на сложност и дублиране. В резултат на това, вместо да опростят процесите, компаниите се оказват в още по-фрагментирана и хаотична информационна среда.



**Фиг. 2.1-7 Сложността на системите и разнообразието от формати на данни водят до загуба на последователност в строителния процес.**

Всички тези проблеми, свързани с управлението на множество разнородни решения, рано или късно довеждат ръководството на компанията до важно осъзнаване: проблемът не е в обема на данните или в търсенето на следващия универсален инструмент за обработка на данни. Истинската причина се крие в качеството на данните и в начина, по който организацията ги създава, получава, съхранява и използва.

Ключът към устойчив успех не е в преследването на нови "вълшебни" приложения, а в изграждането на култура на данните в компанията. Това означава данните да се разглеждат като стратегически актив и качеството, целостта и актуалността на данните да се превърнат в приоритет на всички нива в организацията.

Решението на дилемата "качество срещу количество" се крие в създаването на унифицирана структура на данните, която премахва дублирането, премахва несъответствията и унифицира информационните потоци. Тази архитектура осигурява единен, надежден източник на данни, въз основа на който да се вземат информирани, точни и навременни решения.

В противен случай, както все още често се случва, компаниите продължават да разчитат на субективните мнения и интуитивните оценки на експертите на HiPPO, а не на надеждни факти. В строителния бранш, където експертизата традиционно играе значителна роля, това е особено забележимо.

## HiPPO или опасността от мнения при вземането на решения

Традиционно в строителния бранш ключовите решения се вземат въз основа на опит и субективна преценка. Без навременни и надеждни данни мениджърите на компаниите трябва да действат на сляпо, като разчитат на интуицията на най-високоплатените служители (HiPPO - Highest Paid Person's Opinion), а не на обективни факти (фиг. 2.1-8).

NO ANALYTICS?  
WELCOME TO THE HIPPO\*

\*HIGHEST PAID PERSON'S OPINION



**Фигура 2.1-8 При липсата на анализи бизнесът зависи от субективното мнение на опитни професионалисти.**

Този подход може да е оправдан в стабилна и бавно променяща се среда, но в ерата на цифровата трансформация той се превръща в сериозен рисков. Решенията, основани на интуиция и догадки, са податливи на изкривяване, често се основават на неподкрепени хипотези и не отчитат сложната картина, отразена в данните.

Това, което се приема за интелигентен дебат на ниво вземане на решения в дадена компания, често не се основава на нищо конкретно. Успехът на една компания не трябва да зависи от авторитета и заплатите на експертите, а от способността за ефективна работа с данни, идентифициране на закономерности и вземане на информирани решения.

Важно е да се откажем от схващането, че авторитетът или опитът автоматично означават, че дадено решение е правилно. Подходът, основан на данни, променя правилата на играта: данните и анализите, а не длъжността и заплатата, вече са в основата на вземането на решения. Големите данни, машинното обучение и визуалният анализ ни дават възможност да идентифицираме модели и да разчитаме на факти, а не на догадки (фиг. 1.1-4).

*Без данни вие сте просто още един човек с мнение [34].*

- У. Едуардс Деминг, учен и консултант по управление

Съвременните методи за управление на данни осигуряват и непрекъснатост на знанията в рамките на компанията. Ясно описаните процеси, автоматизацията и систематичният подход позволяват прехвърлянето дори на ключови роли без загуба на ефективност.

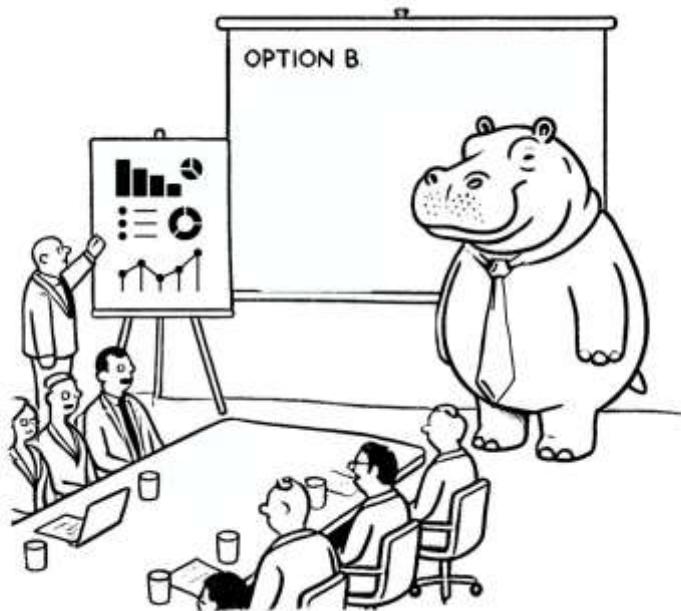
Сляпото доверие в данните обаче също може да доведе до сериозни грешки. Данните сами по себе си са просто сбор от числа. Без подходящ анализ, контекст и способност за идентифициране на модели те нямат стойност и не могат да управляват процесите. Ключът към успеха се крие не в

избора между интуицията на HiPPO и анализа, а в изграждането на интелигентни инструменти, които превръщат разнородната информация в управляеми, информирани решения.

В цифровата среда на строителството решаващи фактори за успех са не старшинството и мястото в йерархията, а бързината на реакция, точността на решенията и ефективността на ресурсите

Данните са инструменти, а не абсолютни истини. Те трябва да допълват човешкото мислене, а не да го заместват. Въпреки ползите от анализите, данните не могат да изместят напълно човешката интуиция и опит. Тяхната роля е да помогат за вземането на по-точни и информирани решения.

Конкурентното предимство ще бъде постигнато не само чрез спазване на стандартите, но и чрез способността да се надминат конкурентите в ефективното използване на ресурсите, които са еднакви за всички. В бъдеще уменията за работа с данни ще станат толкова важни, колкото някога са били грамотността или математическите умения. Специалистите, които могат да анализират и тълкуват данни, ще могат да вземат по-точни решения, измествайки тези, които разчитат само на личния си опит (фигура 2.1-9).



**Фиг. 2.1-9 Решенията трябва да се основават на обективен анализ, а не на мнението на най-високоплатения служител.**

Мениджърите, специалистите и инженерите ще работят като анализатори на данни на , като проучват структурата, динамиката и ключовите показатели на проектите. Човешките ресурси ще се превърнат в елементи на системата, изискващи гъвкаво адаптиране, базирано на данни, за да се постигне максимална ефективност.

*Грешките при използване на неадекватни данни са много по-малки, отколкото при използване на никакви данни [35].*

- Чарлз Бабидж, изобретател на първата аналитична изчислителна машина

Появата на големите данни и въвеждането на LLM (Large Language Models) промениха радикално не само начина на анализиране, но и самия характер на вземането на решения. Докато преди фокусът беше върху причинно-следствената връзка (защо нещо се е случило - диагностичен анализ) (фиг. 1.1-4), днес на преден план излиза възможността за прогнозиране на бъдещето (прогностичен анализ), а в бъдеще и за предписващ анализ, при който машинното обучение и изкуственият интелект предлагат най-добраия избор в процеса на вземане на решения.

Според новото проучване на SAP™, "New Study Finds Nearly Half of Executives Trust Artificial Intelligence More Than Themselves" 2025 [36], 44% от висшите ръководители биха били готови да променят свое предишно решение въз основа на съвет от ИИ, а 38% биха се доверили на ИИ да взема бизнес решения от тяхно име. Същевременно 74% от ръководителите заявяват, че се доверяват на съветите на ИИ повече, отколкото на приятелите и семейството си, а 55% работят в компании, в които прозренията, получени от ИИ, заместват или често заобикалят традиционните методи за вземане на решения - особено в организации с годишни приходи над 5 млрд. долара. Освен това 48% от анкетираните използват генеративни инструменти на ИИ ежедневно, включително 15%, които ги използват по няколко пъти на ден.

С развитието на LLM и автоматизираните системи за управление на данни възниква ново предизвикателство: как да се използва информацията ефективно, без да се губи нейната стойност в хаоса от несъвместими формати и разнородни източници, който се допълва от нарастващата сложност и динамика на бизнес процесите.

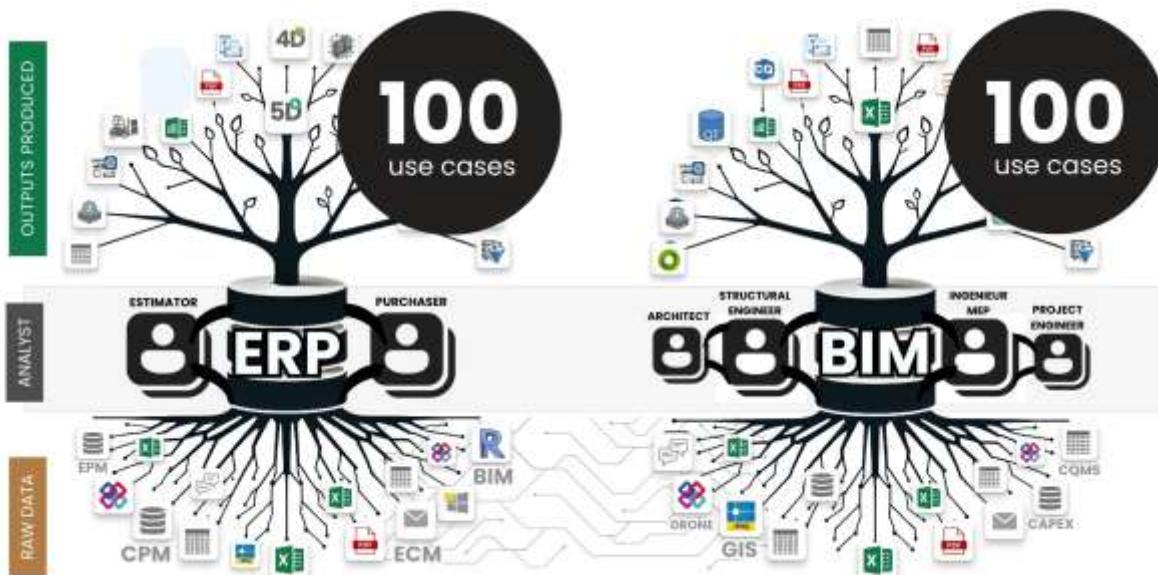
## **Непрекъснато нарастване на сложността и динамиката на бизнес процесите**

Днес строителната индустрия е изправена пред сериозни предизвикателства в управлението на данни и процеси. Основните предизвикателства са свързани с изолираните информационни системи, прекомерната бюрокрация и липсата на интеграция между цифровите инструменти. Тези предизвикателства се засилват, тъй като самите бизнес процеси стават все по-сложни - движени от технологиите, променящите се изисквания на клиентите и развиващите се разпоредби.

Уникалността на строителните проекти се дължи не само на техническите им особености, но и на различията в националните стандарти и нормативните изисквания в различните държави (фиг. 4.2-10, фиг. 5.1-7). Това изиска гъвкав, индивидуален подход към всеки проект, който е труден за изпълнение в рамките на традиционните модулни системи за управление. Поради сложността на процесите и големия обем данни много компании се обръщат към доставчици, предлагачи специализирани решения. Но пазарът е претоварен - много стартиращи компании предлагат сходни продукти, като се фокусират върху тесни задачи. В резултат на това често се губи цялостният подход към управлението на данните.

Адаптирането към непрекъснатия поток от нови технологии и пазарни изисквания се превръща в решаващ фактор за конкурентоспособността. Съществуващите патентовани приложения и модулни системи обаче са с ниска степен на адаптивност - вся-какви промени често изискват продължителни и скъпи преработки от страна на разработчици, които невинаги разбират спецификата на строителните процеси.

Фирмите се оказват заложници на технологичното изоставане, чакайки нови актуализации, вместо бързо да прилагат иновативни интегрирани подходи. В резултат на това вътрешната структура на строителните организации често представлява сложна екосистема от взаимосвързани йерархични и често затворени системи, координирани чрез многостепенна мрежа от мениджъри (фигура 2.1-10).

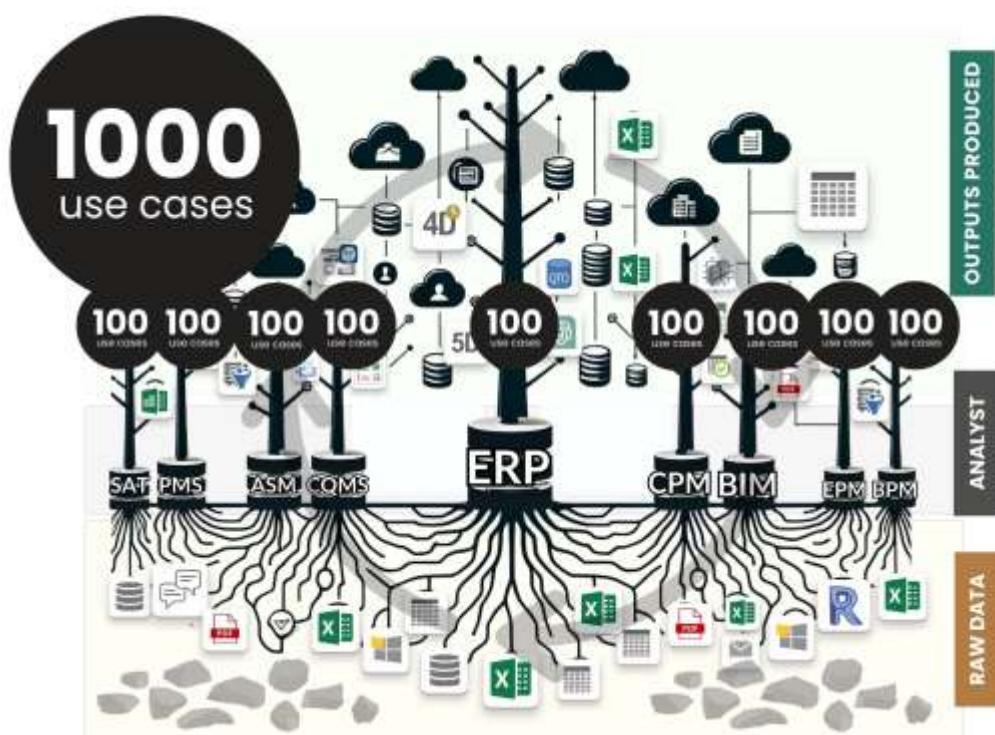


**Фиг. 2.1-10** Компаниите се състоят от взаимосвързани системи, чиято взаимовръзка формира процеси, които изискват автоматизация.

Според проучване, проведено от Канадската строителна асоциация и KPMG Канада през 2021 г. [37], само 25% от компаниите смятат, че са в значителна или различна позиция в сравнение с конкурентите, когато става въпрос за внедряване на технологии или цифрови решения. Само 23% от анкетираните съобщават, че техните решения са значително или в голяма степен базирани на данни. В същото време по-голямата част от участниците в проучването характеризират използването на редица други технологии като чисто експериментално или признават, че изобщо не ги използват.

Това нежелание да се участва в технологични експерименти е особено очевидно при големи инфраструктурни проекти, където грешките могат да струват милиони долари. Дори и най-modерните технологии - цифрови близнаци, прогнозни анализи - често срещат съпротива не заради своята ефективност, а заради липсата на доказана надеждност в реални проекти.

Според доклада на Световния икономически форум (СИФ) "Формиране на бъдещето на строителството" [5] въвеждането на нови технологии в строителството се сблъска не само с технически трудности, но и с психологически бариери от страна на клиентите. [5], въвеждането на нови технологии в строителството се сблъска не само с технически трудности, но и с психологически бариери от страна на клиентите. Много клиенти се опасяват, че използването на съвременни решения ще превърне проектите им в експериментална площадка и ще ги направи "морски свинчета", а непредвидимите последици могат да доведат до допълнителни разходи и рискове.



Фиг. 2.1-11 За всеки случай на използване на данни пазарът на решения предлага приложения за оптимизиране и автоматизиране на процесите.

Строителният бранш е много разнообразен: различните проекти имат различни изисквания, регионални особености, законови разпоредби за класификация (фиг. 4.2-10), изчислителни стандарти (фиг. 5.1-7) и др. Поради това е практически невъзможно да се създаде патентовано универсално приложение или система, която да отговаря напълно на всички тези изисквания и специфики на проекта.

В опит да се справим с нарастващата сложност на системите и зависимостта от доставчиците на софтуер, все повече се осъзнава, че ключът към ефективното управление на данни е не само отвореността и стандартизацията, но и опростяването на самата архитектура на процеса. Нарастващата сложност и динамика на бизнес процесите изисква нови подходи, при които приоритетът се измества от натрупването на данни към тяхното структуриране и организиране. Именно тази промяна ще бъде следващата стъпка в развитието на строителния бранш, която ще отбележи края на

ерата на господството на софтуерните доставчици и началото на ерата на смислената организация на информацията.

Осъзнаването на ограниченията на универсалните решения и уязвимостта към нарастващата сложност води до промяна на приоритетите от затворени платформи и трупане на данни към прозрачност, адаптивност и структурирана обработка на информацията. Тази промяна в мисленето отразява по-широките промени в световната икономика и технологиите, описани през призмата на така наречените "индустриални революции". За да разберем накъде върви строителството и каква е неговата бъдеща посока, е необходимо да разгледаме мястото на отрасъла в контекста на Четвъртата и Петата индустриски революции - от автоматизацията и цифровизацията до персонализацията, отворените стандарти и модела на данни, базиран на услугите.

## **Четвъртата индустриска революция (Индустрия 4.0) и Петата индустриска революция (Индустрия 5.0) в строителството**

Технологичните и икономическите етапи са теоретични концепции, използвани за описание и анализ на еволюцията на обществото и икономиката на различни етапи на развитие. Те могат да бъдат интерпретирани по различен начин от различните изследователи и експерти.

- **Четвъртата индустриска революция (4IR или Индустрис 4.0)** е свързана с информационните технологии, автоматизацията, цифровизацията и глобализацията. Един от ключовите елементи е създаването на собствени софтуерни решения, т.е. специализирани цифрови продукти, предназначени за конкретни задачи и компании. Тези решения често се превръщат във важна част от ИТ инфраструктурата, но са слабо мащабируеми без допълнителни модификации.
- **Петата индустриска революция (5IR)** сега е на по-ранен етап на концептуализация и развитие от 4IR. Нейните основни принципи включват по-голямо персонализиране на продуктите и услугите. 5IR е движение към по-приспособима, гъвкава и персонализирана икономическа дейност с акцент върху персонализацията, консултациите и моделите, ориентирани към услугите. Ключов аспект на петия икономически режим е използването на данни за вземане на решения, което е практически невъзможно без използването на отворени данни и отворени инструменти (фигура 2.1-12).

FOURTH INDUSTRIAL REVOLUTION  
(4IR OR INDUSTRY 4.0)



FIFTH INDUSTRIAL REVOLUTION  
(5IR)



Фиг. 2.1-12 Четвъртият модел се фокусира върху решенията, докато петият модел се фокусира върху персонализацията и данните.

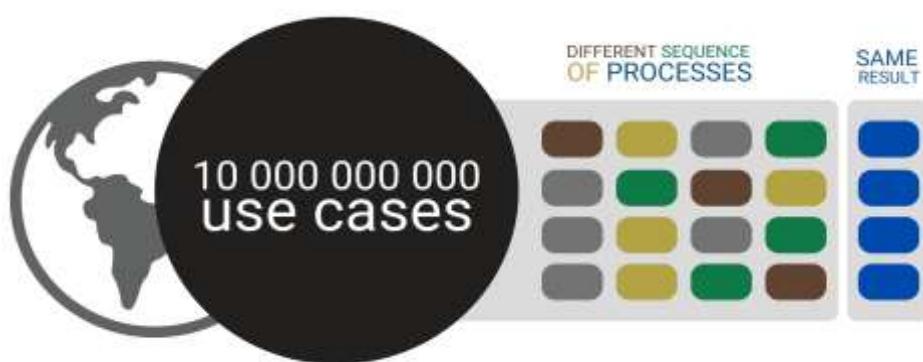
Създаването на приложение за фирми от строителния бранш, което да се използва в десет или сто организации, не гарантира успешното му разпространение в други фирми, региони или държави без значителни промени и подобрения. Вероятността за успешно мащабиране на такива решения остава малка, тъй като всяка организация има уникални процеси, изисквания и условия, които могат да изискват персонализирани адаптации.

Важно е да се разбере, че още днес успешното интегриране на технологични решения предполага дълбоко персонализиран подход към всеки процес, проект и компания. Това означава, че дори след като е разработена универсална рамка, инструмент или програма, тя ще изиска детайлно адаптиране и персонализиране, за да отговори на уникалните изисквания и условия на всяка конкретна компания и проект.

Според доклада на PwC "Декодиране на петата индустриална революция" [38] [38], около 50% от висшите ръководители в различни индустрии през тази година разчитат на интегрирането на напреднали технологии и човешки опит. Този подход им позволява бързо да се адаптират към промените в дизайна на продуктите или изискванията на клиентите, създавайки персонализирано производство.

Всеки процес изиска разработване на уникална функция или приложение, което, предвид размера на световната строителна индустрия и разнообразието на проектите, води до съществуването на огромен брой бизнес казуси, представляващи всеки път уникална логика на тръбопровода (фиг. 2.1-13). Всеки такъв случай има своите особености и изиска индивидуален подход. Ще разгледаме по-подробно разнообразието от възможни решения на един и същ аналитичен проблем в контекста на различни подходи в главата, посветена на машинното обучение и анализа на набора от данни за "Титаник" (фиг. 9.2-9).

**Конвейер** в контекста на цифровите процеси е последователност от дейности, процеси и инструменти, които позволяват автоматизиран или структуриран поток от данни и работа през различните етапи от жизнения цикъл на проекта.



Фиг. 2.1-13 Индивидуалността и променливостта на бизнес казусите прави невъзможни опитите за създаване на мащабиращи затворени платформи и инструменти.

Животът ни вече се промени в много отношения под влиянието на цифровата трансформация, а днес можем да говорим за настъпването на нов етап в икономическото развитие на строителната индустрия. В тази "нова икономика" конкуренцията ще бъде организирана по други правила: този, който е в състояние ефективно да превърне публичните знания и отворените данни в търсени продукти и услуги, получава ключово предимство в условията на петата индустриална революция.

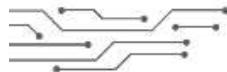
Както отбелязва икономистът Кейт Маскус в книгата "Частни права и обществени проблеми: глобалната икономика на интелектуалната собственост през 21-ви век" [39]. 2012 г. [39], "живеем в глобална икономика на знанието и бъдещето принадлежи на онези, които знаят как да превърнат научните открития в стоки".

Преходът към петия икономически режим предполага промяна на фокуса от затворени ИТ решения към отворени стандарти и платформи. Компаниите ще започнат да се отказват от традиционните софтуерни продукти в полза на модели, ориентирани към услугите, при които данните, а не патентованите технологии, ще се превърнат в основен актив.

Проучване на Harvard Business School 2024 [40] показва огромната икономическа стойност на софтуера с отворен код (Open Source Software, OSS). Според проучването OSS присъства в 96% от всички софтуерни кодове, а някои комерсиални софтуери се състоят от 99,9% OSS компоненти. Без OSS компаниите биха изразходвали 3,5 пъти повече средства за софтуер.

Изграждането на фирмени екосистеми, следвайки световните тенденции, постепенно ще премине към пета икономическа парадигма, в която ориентираните към данните анализи и консултантски услуги ще станат по-приоритетни от изолираните, затворени решения с твърдо определени сценарии на използване.

Ерата на цифровизацията ще промени баланса на силите в индустрията: вместо да разчитат на решения на доставчиците, компаниите ще основават конкурентоспособността си на способността си да използват ефективно данните. В резултат на това строителният бранш ще премине от наследени закостенели системи към гъвкави, адаптивни екосистеми, в които отворените стандарти и оперативно съвместимите инструменти ще бъдат в основата на управлението на проекти. Краят на ерата на доминиране на доставчиците на приложения ще създаде нова среда, в която стойността ще се определя не от притежаването на затворен код и специализирани конектори, а от способността да се превърнат данните в стратегическо предимство.



## ГЛАВА 2.2.

### ПРЕВРЪЩАНЕ НА ХАОСА В РЕД И НАМАЛЯВАНЕ НА СЛОЖНОСТТА

#### **Излишният код и затворените системи като пречка за подобряване на производителността**

През последните десетилетия технологичните промени в сферата на ИТ се движат предимно от доставчиците на софтуер. Те определят посоката на развитие, като определят кои технологии да бъдат възприети от компаниите и кои да бъдат изоставени. В епохата на преминаване от изолирани решения към централизирани бази данни и интегрирани системи доставчиците популяризираха лицензиирани продукти, осигуряващи контрол върху достъпа и мащабируемостта. По-късно, с навлизането на облачните технологии и моделите "софтуер като услуга" (SaaS), този контрол еволюира в абонаментен модел, затвърждавайки потребителите като лоялни клиенти на цифрови услуги.

Този подход доведе до парадокс: въпреки безprecedентните обеми на създадения програмен код, само малка част от него се използва в действителност. Може би има стотици или хиляди пъти повече код от необходимото, тъй като едни и същи бизнес процеси се описват и дублират в десетки или стотици програми по различни начини, дори в рамките на една и съща компания. В същото време разходите за разработване вече са платени, а тези разходи са невъзстановими. Въпреки това индустрията продължава да възпроизвежда този цикъл, създавайки нови продукти с минимална добавена стойност за крайния потребител, по-често под натиска на пазарните очаквания, отколкото на реалните нужди.

Според Ръководството за оценка на разходите за разработване на софтуер на Университета за отбранителна аквизиция (DAU) [41] разходите за разработване на софтуер могат да варират значително в зависимост от няколко фактора, включително сложността на системата и избраната технология. Исторически погледнато, разходите за разработка за 2008 г. са около 100 USD на ред изходен код (SLOC), докато разходите за поддръжка могат да достигнат до 4000 USD на SLOC.

Само един от компонентите на CAD приложението - геометричното ядро - може да съдържа десетки милиони редове код (фиг. 6.1-5). Подобна ситуация се наблюдава и при ERP системите (фиг. 5.4-4), към обсъждането на чиято сложност ще се върнем в петата част на книгата. При по- внимателно вглеждане обаче се оказва, че голяма част от този код не добавя стойност, а просто действа като "пощальон" - механично пренася данни между базата данни, API, потребителския интерфейс и други таблици в системата. Въпреки популярния мит за критичната важност на т. нар. бизнес логика, сировата реалност е много по-прозаична: съвременните бази от кодове са пълни с оstarели шаблонни блокове (legacy code), чиято единствена цел е да осигурят прехраняването на данни между таблици и компоненти, без това да се отрази на вземането на решения или на ефективността на бизнеса.

В резултат на това затворените решения, които обработват данни от различни източници, неизбежно се превръщат в объркващи "екосистеми от спагети". Тези сложни, преплетени системи могат да бъдат управлявани само от армия от мениджъри, работещи в полурутинен режим. Тази организация на управление на данните е не само неефективна от гледна точка на ресурсите, но и

създава критични уязвимости в бизнес процесите, като прави компанията зависима от тесен кръг специалисти, които разбират как работи този технологичен лабиринт.

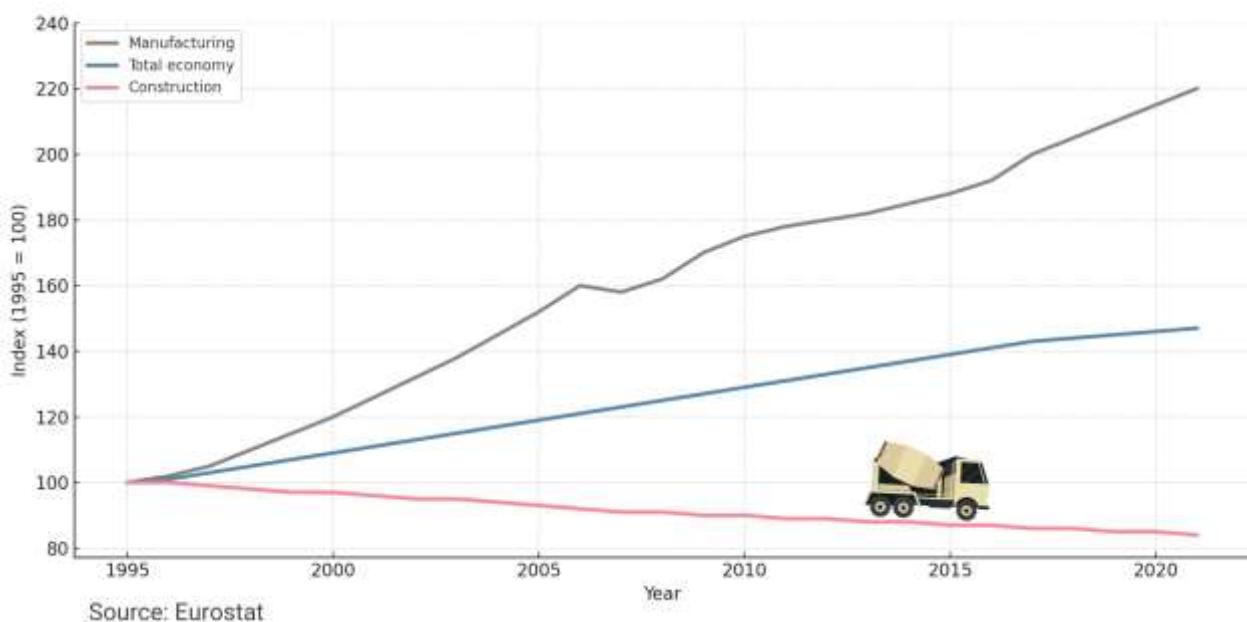
Непрекъснатото увеличаване на обема на кода, броя на приложенията и нарастващата сложност на концепциите, предлагани от доставчиците, доведе до естествен резултат - увеличаване на сложността на ИТ екосистемата в строителството. Това направи неефективно практическото прилагане на цифровизацията чрез увеличаване на броя на приложенията в отрасъла. Софтуерните продукти, създадени без необходимото внимание към нуждите на потребителите, често изискват значителни ресурси за внедряване и поддръжка, но не носят очакваната възвръщаемост.

Според проучването на McKinsey "Повишаване на производителността в строителството" [42], през последните две десетилетия глобалният ръст на производителността на труда в строителството е средно само 1% годишно в сравнение с ръст от 2,8% за световната икономика като цяло и 3,6% за промишлеността. В Съединените щати производителността на труда на работник в строителството е намаляла наполовина от 60-те години на миналия век [43].

Нарастващата сложност на системите, изолацията и затворените данни влошават комуникацията между специалистите, което прави строителната индустрия една от най-неефективните (фигура 2.2-1). до 22 трилиона долара до 2040 г., което ще изисква значително повишаване на ефективността.

**Productivity growth in Manufacturing, Total Economy, and Construction (1995-2021)**

Labour productivity of the added value in volume per hour worked in the EU, 1995-2021 (index 1995=100)



**Фиг. 2.2-1** Затворените и сложни данни и като следствие от това лошата комуникация между специалистите доведоха строителната индустрия до един от най-неефективните сектори на икономиката (по [44], [45]).

Както се подчертава в проучването на McKinsey (2024 г.) "Осигуряването на производителност в строителството вече не е по избор", с нарастващия недостиг на ресурси и стремежа на отрасъла да удвои темпа си на растеж, строителството вече не може да си позволи да остане на сегашните нива на производителност [44]. Предвижда се глобалните разходи за строителство да нараснат от 13 трилиона долара през 2023 г. до много по-високи нива до края на десетилетието, което прави въпроса за ефективността не просто актуален, а критичен.

Един от основните начини за подобряване на ефективността ще бъде неизбежното унифициране и опростяване на структурите на приложенията и архитектурите на екосистемите за данни. Този подход към рационализацията ще премахне излишните слоеве на абстракция и ненужната сложност, натрупани през годините в системите на предприятията.

## От отделни системи към единен склад за данни

Колкото повече данни натрупва една организация, толкова по-трудно става извлечането на реална стойност от тях. Поради фрагментарния характер на съхраняването на информацията в изолирани силози, бизнес процесите на съвременните компании приличат на строители, които се опитват да построят небостъргач от материали, съхранявани в хиляди различни складове. Излишъкът от информация не само затруднява достъпа до правно релевантна информация, но и забавя вземането на решения: всяка стъпка трябва да се проверява и потвърждава многократно.

Всяка задача или процес са свързани с отделна таблица или база данни, а обменът на данни между системите изисква сложни интеграции. Грешките и несъответствията в една система могат да доведат до верижни сривове в други. Неправилните стойности, закъслелите актуализации и дублиращата се информация принуждават служителите да прекарват значително време в ръчно съгласуване и коригиране на данните. В резултат на това организацията отделя повече време за справяне с последиците от фрагментацията, отколкото за разработване и оптимизиране на процесите.

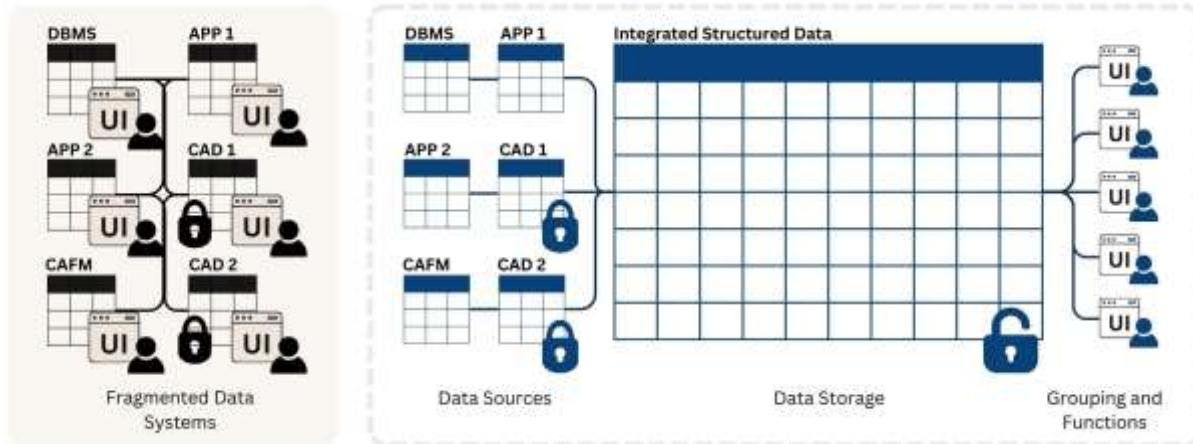
Този проблем е универсален: някои компании продължават да се борят с хаоса, докато други намират решение в интеграцията - преместване на информационните потоци в централизирана система за съхранение. Мислете за нея като за една голяма таблица, в която можете да съхранявате всяка същност, свързани със задачи, проекти и обекти. Вместо десетки разнородни таблици и формати се появява едно цялостно хранилище (фигура 2.2-2), което позволява

- да сведете до минимум загубата на данни;
- да се премахне необходимостта от постоянно хармонизиране на информацията;
- подобряване на наличността и качеството на данните;
- опростяване на аналитичната обработка и машинното обучение

Привеждането на данните в общ стандарт означава, че независимо от източника, информацията се преобразува в унифициран и машинно четим формат. Подобна организация на данните позволява да се проверява тяхната цялост, да се анализират в реално време и да се използват своевременно за вземане на управленски решения.

Концепцията за интегрирани системи за съхранение и тяхното приложение в анализа и машинното

обучение ще бъдат разгледани по-подробно в главата "Съхранение на големи обеми данни и машинно обучение". Темите за моделиране и структуриране на данни ще бъдат разгледани подробно в главите "Трансформиране на данни в структурирана форма" и "Как стандартите променят играта: от случайни файлове към сложен модел на данните".



**Фиг. 2.2-2 Интеграцията на данни премахва "силозите", подобрява наличността на информация и оптимизира бизнес процесите.**

След като данните са структурирани и обединени, следващата логична стъпка е тяхното валидиране. С едно интегрирано хранилище този процес е значително опростен: вече няма множество непоследователни схеми, дублиращи се структури и сложни връзки между таблиците. Цялата информация е приведена в съответствие с единен модел на данните, което елиминира вътрешните несъответствия и ускорява процеса на валидиране. Валидирането и осигуряването на качеството на данните са крайъгълни аспекти на всички бизнес процеси и ще ги разгледаме по-подробно в съответните глави на книгата.

На последния етап данните се групират, филтрират и анализират. Към тях се прилагат различни функции: агрегиране (събиране, умножаване), изчисления между таблици, колони или редове (фиг. 2.2-4). Работата с данни се превръща в последователност от стъпки: събиране, структуриране, валидиране, трансформиране, аналитична обработка и разтоварване към крайните приложения, където информацията се използва за решаване на практически проблеми. Ще обсъдим повече за изграждането на такива сценарии, автоматизирането на стъпките и изграждането на потоци за обработка в главите за ETL -процесите и подхода на тръбопроводите за данни.

Така че цифровата трансформация не се състои само в опростяване на обработката на информация. Става дума за премахване на прекомерната сложност при управлението на данните, за преминаване от хаос към предвидимост, от множество системи към управляем процес. Колкото по-ниска е сложността на архитектурата, толкова по-малко код е необходим за нейната поддръжка. А в бъдеще кодът като такъв може да изчезне напълно, отстъпвайки място на интелигентни агенти, които самостоятелно анализират, систематизират и трансформират данните.

## Интегрираните системи за съхранение позволяват преминаването към агенти с изкуствен интелект

Колкото по-малко сложни са данните и системите, толкова по-малко код трябва да се пише и поддържа. А най-лесният начин за спестяване на средства за разработка е да се отървete изцяло от кода, като го замените с данни. Когато разработката на приложения преминава от код към модели на данни, неминуемо се преминава към подход, ориентиран към данните (data-driven), защото зад тези концепции стои напълно различен начин на мислене.

Когато човек избере да работи с данни в центъра, той започва да вижда ролята им по различен начин. Данните вече не са просто "сировина" за приложениета - те вече са основата, около която се изграждат архитектурата, логиката и взаимодействието.

Традиционният подход към управлението на данните обикновено започва на ниво приложение и в строителството наподобява тромава бюрократична система: многостепенни одобрения, ръчни проверки, безкрайни версии на документи чрез съответните софтуерни продукти. С развитието на цифровите технологии все повече компании ще бъдат принудени да преминат към принципа на минимализма - да съхраняват и използват само това, което наистина е необходимо и ще бъде използвано.

Логиката на минимизиране е възприета от продавачите. За да се опрости съхранението и обработката на данни, работата на потребителите се прехвърля от офлайн приложения и инструменти към облачни услуги и така наречените SaaS решения.

Концепцията SaaS (Software as a Service, или "софтуер като услуга") е една от основните тенденции в съвременните ИТ инфраструктури, която позволява на потребителите да имат достъп до приложения чрез интернет, без да се налага да инсталират и поддържат софтуер на собствените си компютри.

От една страна, SaaS улесни мащабирането, контрола на версийте и намали разходите за поддръжка и обслужване, но от друга страна, освен зависимостта от логиката на дадено приложение, направи потребителя напълно зависим от облачната инфраструктура на доставчика. Ако дадена услуга се срине, достъпът до данни и бизнес процеси може да бъде временно или дори постоянно блокиран. Освен това всички потребителски данни при работа с приложения SaaS се съхраняват на сървърите на доставчика, което създава рискове за сигурността и спазването на нормативните изисквания. Промените в тарифите или условията за ползване също могат да доведат до увеличаване на разходите или до необходимост от спешна миграция.

Разработването на AI, LLM -агенти и подход, ориентиран към данните, постави под въпрос бъдещето на приложениета в традиционната им форма и SaaS изпълнение. Докато преди се изискваше приложениета и услугите да управляват бизнес логиката и да обработват данни, с появата на агентите с ИИ тези функции могат да се прехвърлят към интелигентни системи, които работят директно с данни.

Ето защо в ИТ отделите и на управлениско ниво все по-често се обсъждат хибридни архитектури, при които агентите с изкуствен интелект и локалните решения допълват облачните услуги, като

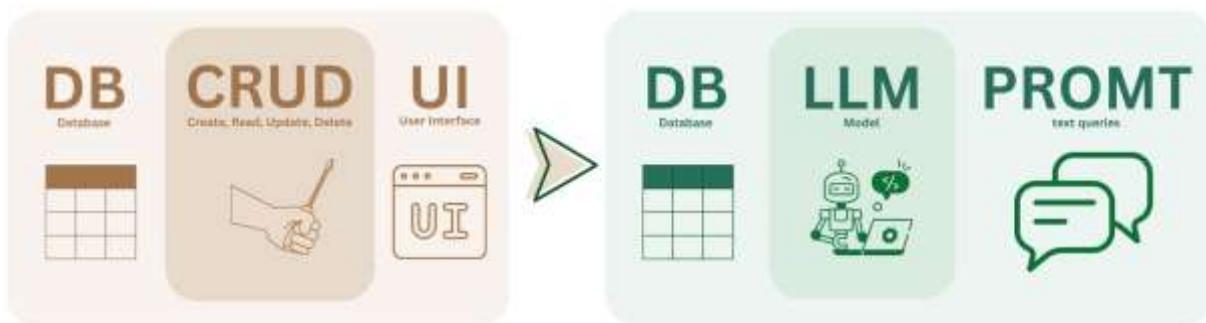
намаляват зависимостта от SaaS -платформи.

*Подходът, който прилагаме, отчита, че традиционните бизнес приложения или приложенията SaaS могат да се променят драстично в ерата на агентите. Тези приложения по същество са CRUD [създаване, четене, актуализиране и изтриване] бази данни с бизнес логика. Но в бъдеще тази логика ще бъде поета от агенти с изкуствен интелект [46].*

- Сатя Надела, главен изпълнителен директор на Microsoft, 2024 г.

Подходът, ориентиран към данните, и използването на агенти с изкуствен интелект и управление на личния живот могат да намалят излишните процеси и по този начин да намалят работното на товарване на служителите. Когато данните са организирани правилно, става по-лесно да се анализират, визуализират и прилагат при вземането на решения. Вместо безкрайни отчети и проверки, специалистите получават достъп до актуална информация с няколко кликвания или с помощта на агенти на LLM автоматично под формата на готови документи и информационни табла.

В манипулирането на данните ще ни помогат инструментите за изкуствен интелект (AI) и чатовете LLM. През последните години се наблюдава тенденция към отказ от традиционните CRUD операции (създаване, четене, актуализиране, изтриване) към използване на големи езикови модели (LLM) за управление на данни. LLM са способни да интерпретират естествен език и автоматично да генерират подходящи заявки към бази данни, което опростява взаимодействието със системите за управление на данни (фиг. 2.2-3).



**Фиг. 2.2-3 ИИ ще замени и интегрира решенията за съхранение и бази данни, като постепенно ще измести традиционните приложения и CRUD -операциите.**

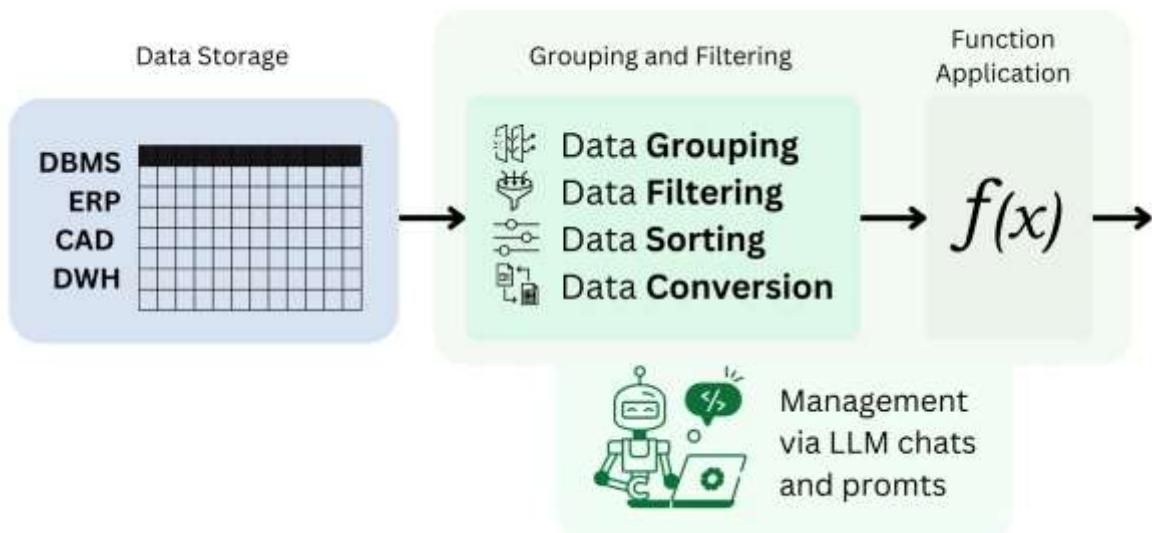
През следващите 3-6 месеца ИИ ще пише 90% от кода, а след 12 месеца почти целият код може да се генерира от ИИ [47].

- Дарио Амодей, главен изпълнителен директор на LLM Anthropic, март 2025 г.

Въпреки бързото развитие на инструментите за разработване на изкуствен интелект (напр. GitHub

Copilot), през 2025 г. разработчиците все още ще играят ключова роля в този процес. Агентите на ИИ се превръщат във все по-полезни асистенти: те автоматично интерпретират заявките на потребителя, генерират заявки на SQL и Pandas (повече за това в следващите глави) или пишат код за анализ на данни. По този начин изкуственият интелект постепенно замества традиционните потребителски интерфейси на приложенията.

Разпространението на модели на изкуствен интелект, като например езикови модели, ще стимулира разработването на хибридни архитектури. Вместо да се откажем напълно от облачните решения и продуктите SaaS, може да станем свидетели на интегриране на облачните услуги с локални системи за управление на данни. Например, обединеното обучение дава възможност за създаване на мощни модели на изкуствен интелект, без да се налага преместване на чувствителни данни в облака. По този начин компаниите могат да запазят контрола върху своите данни, като същевременно получат достъп до усъвършенствани технологии.



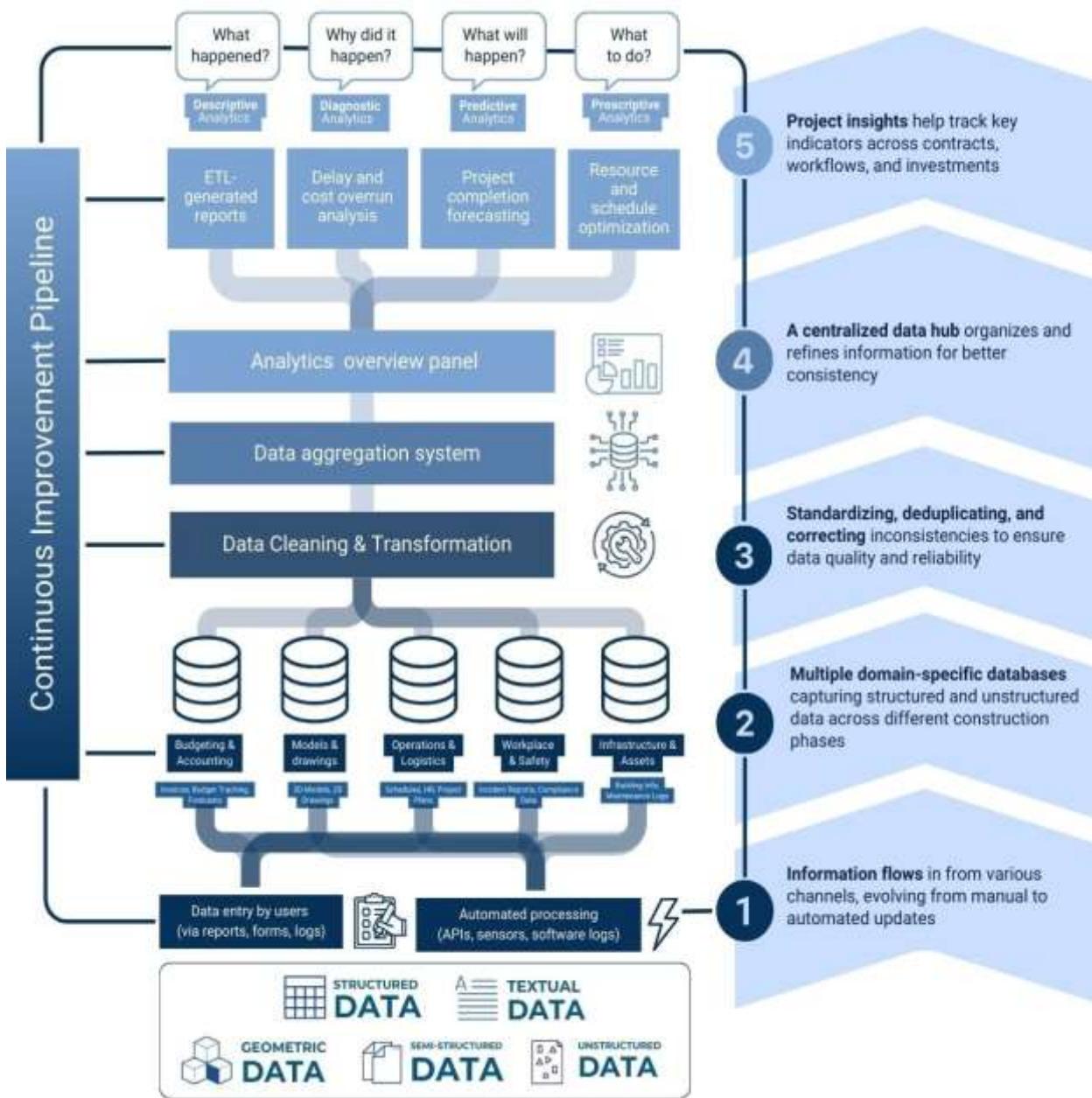
**Фигура 2.2-4 Основните операции по групиране, филтриране и сортиране, последвани от прилагане на функции, ще се обработват от чатовете на LLM.**

Бъдещето на строителната индустрия ще се основава на комбинация от локални решения, облачна мощ и интелигентни модели, които работят заедно за създаване на ефективни и сигурни системи за управление на данни. LLM ще даде възможност на потребителите без задълбочени технически познания да взаимодействат с бази данни и складове за данни, като формулират заявките си на естествен език. Ще разкажем повече за агентите на LLM и AI и как работят те в главата "Агенти на LLM и структурирани формати за данни".

Правилно организираните данни и простите, лесни за използване аналитични инструменти, поддържани от LLM, не само ще улеснят работата с информацията, но и ще помогнат да се сведат до минимум грешките, да се повиши ефективността и да се автоматизират процесите.

## **От събиране на данни до вземане на решения: пътят към автоматизацията**

В следващите части на книгата ще разгледаме подробно как специалистите взаимодействат помежду си и как данните стават основа за вземане на решения, автоматизация и оперативна ефективност. На фигура 2.2-5 е представена примерна диаграма, показваща последователността на стъпките за обработка на данни при подход, ориентиран към данните. Тази диаграма илюстрира тръбопровода за непрекъснато подобрене), части от който ще бъдат разгледани подробно по-нататък в книгата.



Фиг. 2.2-5 Пример за конвейер за непрекъснато подобряване на данните: потокът за обработка и анализ на данни в строителните проекти.

Системата, описваща бизнес процесите на една средна по големина компания, е изградена на принципа на няколко нива. Тя включва: събиране на данни, почистване, обобщаване, аналитична обработка и вземане на решения въз основа на резултатите. Всички тези етапи ще разгледаме по-нататък в книгата - както в теоретичен контекст, така и чрез практически примери:

- На първото ниво се извършва **въвеждане на данни** (фиг. 3.1-1). Информацията се получава както ръчно (чрез отчети, формуляри, дневници), така и в автоматизирана форма (от API, сензори, софтуерни системи). Данните могат да бъдат с различна структура: геометрични, текстови, неструктурирани. На този етап има нужда от стандартизация, структуриране и

унифициране на информационните потоци.

- Следващото ниво е **обработката и преобразуването на данни**. То включва процесите на почистване, отстраняване на дублиращи се данни, коригиране на грешки и подготовка на информацията за по-нататъшен анализ (фигура 4.2-5). Този етап е от решаващо значение, тъй като качеството на анализите пряко зависи от чистотата и точността на данните.
- След това **данныте постъпват в специализирани таблици, рамки за данни или бази данни**, разделени по функционални области: бюджетиране и счетоводство, модели и чертежи, логистика, сигурност и инфраструктура. Това разделение дава възможност за лесен достъп и кръстосан анализ на информацията.
- След това данните се **обобщават и показват в аналитично табло** (витрина). Тук се прилагат описателни, диагностични, прогнозни и прескриптивни методи за анализ. Това помага да се отговори на ключови въпроси (фигура 1.1-4): какво се е случило, защо се е случило, какво ще се случи в бъдеще и какви действия трябва да се предприемат. Например системата може да идентифицира закъснения, да прогнозира завършването на проектите или да оптимизира ресурсите.
- И накрая, последното ниво генерира **аналитични заключения и ключови показатели**, които помагат да се следи изпълнението на договорите, да се управляват инвестициите и да се подобряват бизнес процесите (фиг. 7.4-2). Тази информация се превръща в основа за вземане на решения и за стратегията за развитие на компанията.

По същия начин данните преминават от събиране към използване в стратегическото управление. В следващите части на книгата ще разгледаме подробно всеки етап, като се фокусираме върху видовете данни, техниките за обработка на данни, инструментите за анализ и реални случаи на използване на тези подходи в строителната индустрия.

## Следващи стъпки: превръщане на хаоса в управляема система

В тази част разглеждахме предизвикателствата, свързани с информационните силози, и разгледахме въздействието на прекомерната сложност на системите върху ефективността на бизнеса, като анализирахме прехода от четвъртата индустриална революция към петата, в която централно място заемат данните, а не приложенията. Видяхме как изолираните информационни системи създават пречки за споделяне на знания, а продължаващата сложност на информационния пейзаж намалява производителността и възпрепятства иновациите в строителната индустрия.

За да обобщим тази част, си струва да подчертаем основните практически стъпки, които ще ви помогнат да прилагате обсъжданите подходи в ежедневните си задачи:

- Визуализирайте информационния си пейзаж
  - Създайте визуална карта на източниците на данни (Miro, Figma, Canva), с които работите редовно.
  - Добавете системите и приложенията, които използвате в работата си, към тази карта
  - Идентифициране на потенциално дублиращи се функции и излишни решения
  - Идентифициране на критичните точки, в които може да възникне загуба или повреда на данни по време на предаването им между системите

■ Прилагане на персонализирани практики за управление на данни

- Преместване на фокуса от приложенията към данните като ключов актив в процесите
- Документиране на източниците на данни и методологията за обработка с цел осигуряване на прозрачност
- Разработване на механизми за оценка и подобряване на качеството на данните
- Стремете се да гарантирате, че данните се въвеждат веднъж и се използват много-кратно - това е в основата на ефективната организация на процесите.

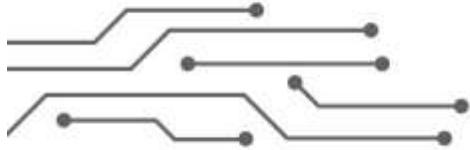
■ Насърчаване на подход, ориентиран към данните (data-driven) във вашия екип

- Предлагат използването на стандартизириани и единни формати за обмен на данни между равнопоставени партньори.
- Редовно повдигайте въпроси, свързани с качеството и наличността на данните, на срещите на екипа.
- Запознайте се с алтернативите с отворен код на инструментите, които използвате за решаване на проблемите си

Започнете с малки проекти - изберете един конкретен процес или набор от данни, който е от решаващо значение за работата ви, и приложете към него подход, ориентиран към данните, като изместите фокуса от инструментите към данните. Постигайки успех в един пилотен проект, ще получите не само практически опит, но и ясна демонстрация на ползите от новата методология за вашия екип. При изпълнението на повечето от тези стъпки, ако имате въпроси, можете да потърсите разяснения и съдействие от всеки актуален LLM.

В следващите части на книгата ще преминем към по-подробен преглед на техниките за структуриране и хармонизиране на данни и ще разгледаме практически подходи за интегриране на разнородна информация. Особено внимание ще бъде обърнато на прехода от разнородни силози към унифицирани екосистеми от данни, които играят ключова роля в цифровата трансформация на строителната индустрия.





### III ЧАСТ

## РАМКА ЗА ДАННИ В БИЗНЕС ПРОЦЕСИТЕ В СТРОИТЕЛСТВОТО

В третата част се формира цялостна представа за типологията на данните в строителството и методите за тяхното ефективно организиране. Анализират се характеристиките и спецификите на работата със структурирани, неструктурни, полуструктурирани, текстови и геометрични данни в контекста на строителните проекти. Разглеждат се съвременните формати за съхранение и протоколи за обмен на информация между различни системи, използвани в отрасъла. Описани са практически инструменти и техники за конвертиране на данни в различни формати в единна структурирана среда, включително как да се интегрират данни от CAD (BIM). Предложени са подходи за осигуряване на качеството на данните от чрез стандартизация и валидиране, което е от решаващо значение за точността на строителните изчисления. Подробно са анализирани практическите аспекти на използването на съвременни технологии (Python Pandas, LLM -модели) с примери за кодове за решаване на типични проблеми в строителната индустрия. Обоснована е стойността на създаването на център за компетентност (CoE) като организационна структура за координация и стандартизация на подходите за управление на информацията.

## ГЛАВА 3.1.

# ТИПОВЕ ДАННИ В СТРОИТЕЛСТВОТО

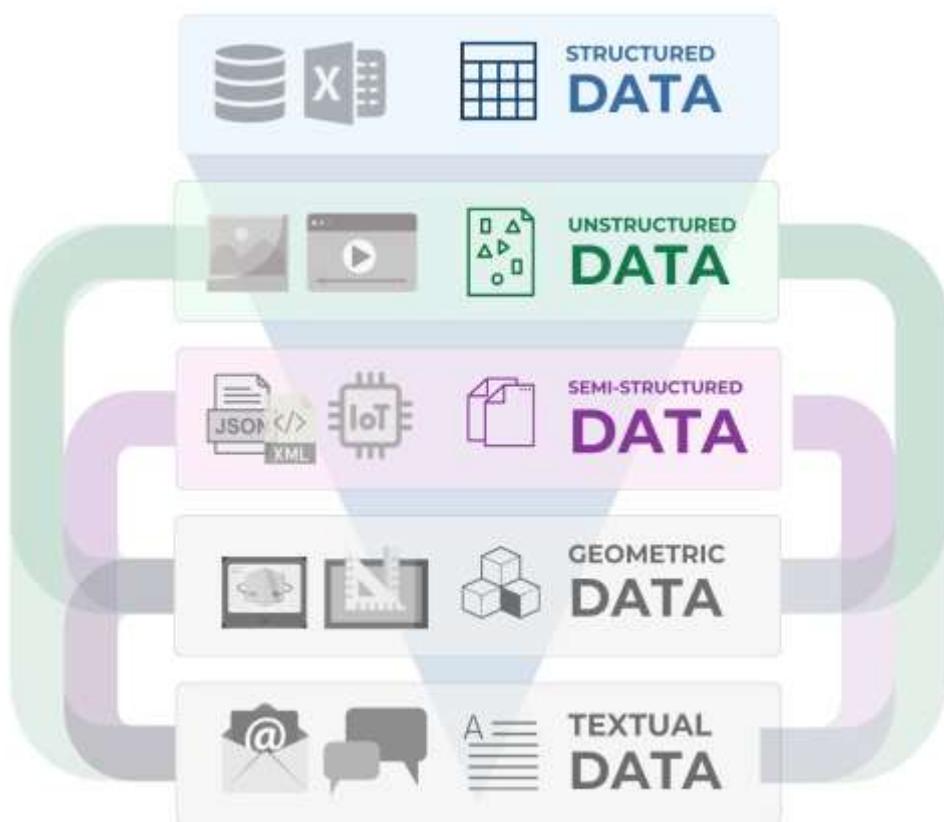
### Най-важните типове данни в строителната индустрия

В съвременната строителна индустрия системите, приложенията и складовете за данни на компаниите активно се изпълват с информация и данни от различни видове и формати (фиг. 3.1-1). Нека разгледаме по-подробно основните видове данни, които формират информационния ландшафт на една съвременна компания, работеща в строителния бранш:

- **Структурирани** данни: тези данни имат ясна организационна структура, напр. електронни таблици на Excel и релационни бази данни.
- **Неструктурирани** данни: това е информация, която не е организирана според строги правила. Примери за такива данни са текст, видео, снимки и аудиозаписи.
- **Свободно структурирани** данни: тези данни заемат междинно положение между структурираните и неструктурните данни. Те съдържат елементи на структура, но тази структура невинаги е ясна или често се описва чрез различни схеми. Примери за полуструктурни данни в строителството са: технически спецификации, проектна документация или доклади за напредъка.
- **Текстови** данни: включва всичко, което произлиза от устни и писмени комуникации, като например имейли, стенограми от срещи и срещи.
- **Геометрични** данни: тези данни идват от CAD програми, в които специалистите създават геометрични данни на елементите на проекта за визуализация, потвърждаване на обемните стойности или проверка на колизии.

Важно е да се отбележи, че геометричните и текстовите (буквено-цифрови) данни не са отделна категория, а могат да присъстват и в трите вида данни. Геометричните данни, например, могат да бъдат част както от структурирани данни (параметрични CAD формати), така и от неструктурни данни (сканирани чертежи). Текстовите данни могат по същия начин да бъдат както организирани в бази данни (структурни данни), така и да съществуват като документи без ясна структура.

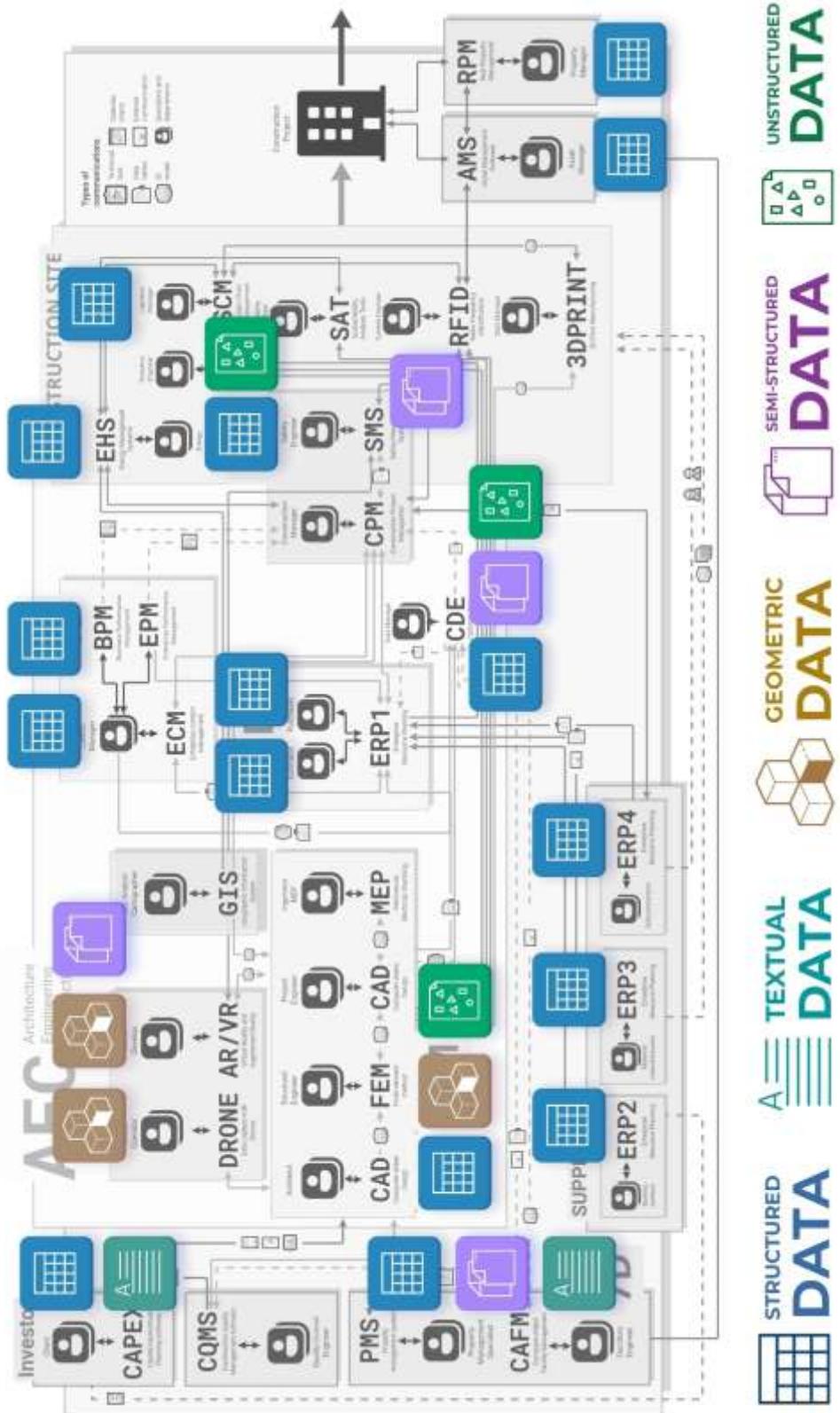
Всеки тип данни в строителната компания е уникален елемент от мозайката на информационните активи на компанията. От неструктурни данни, като например изображения от строителни обекти и аудиозаписи на срещи, до структурирани записи, включително таблици и бази данни, всеки елемент играе важна роля в оформянето на информационния пейзаж на компаниата.



Фиг. 3.1-1 Инженерите и мениджърите на данни трябва да се научат да работят с всички видове данни, използвани в строителната индустрия.

Ето примерен списък само на някои от системите и свързаните с тях типове данни (фигура 3.1-2), използвани в строителството:

- **ERP** (Планиране на ресурсите на предприятието) - обработва общоструктурирани данни, за да подпомогне управлението на ресурсите на предприятието и интегрирането на различни бизнес процеси.
- **CAD** (Компютърно проектиране) в комбинация с **BIM** (Информационно моделиране на сгради) - използва геометрични и полуструктурни данни за проектиране и моделиране на строителни проекти, като осигурява точност и последователност на информацията по време на фазата на проектиране.
- **ГИС** (Географски информационни системи) - работи с геометрични и структурирани данни, за да създава и анализира картографски данни и пространствени връзки.
- **RFID** (Radio-Frequency Identification) - използва полуструктурни данни за ефективно проследяване на материали и оборудване на строителната площадка чрез радиочестотна идентификация.
- **ECM** (Engineering Content Management - управление на инженерно съдържание) е система за управление на инженерни данни и документация, включително полуструктурни и неструктурни данни, като например технически чертежи и документи за проектиране.

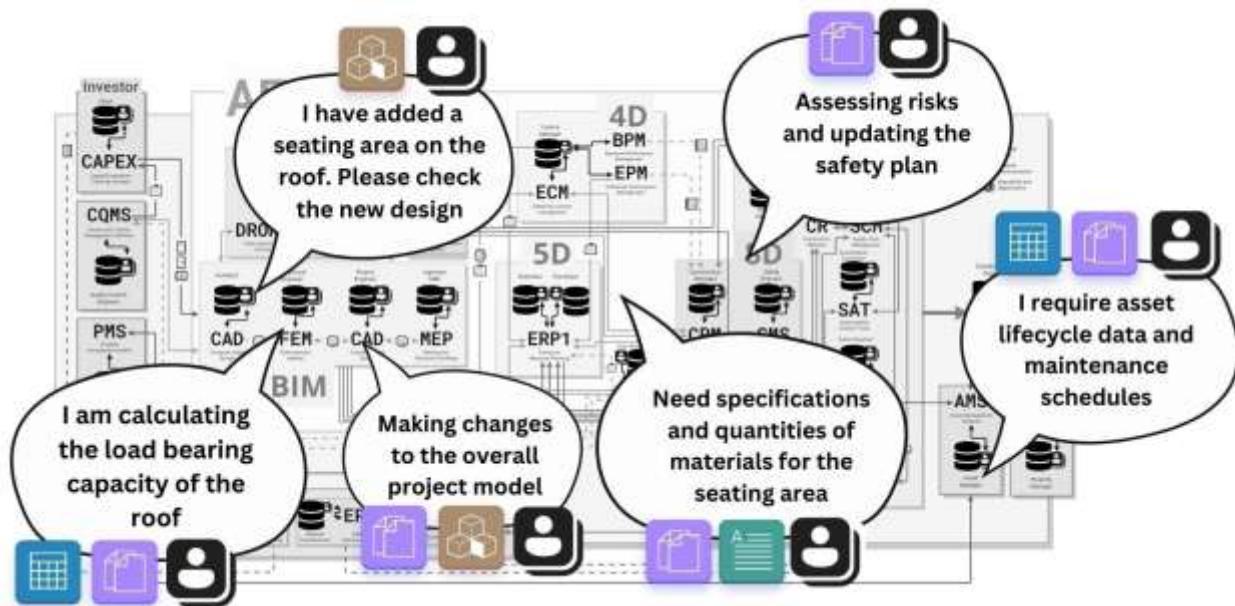


Фиг. 3.1-2 РАЗЛИЧНИ ФОРМАТИ И ДАННИ ПОПЪЛВАТ РАЗЛИЧНИ СИСТЕМИ, КОЕТО ИЗИСКВА ПРЕВОД ВЪВ ФОРМА, ПОДХОДЯЩА ЗА СЛОЖНА ИНТЕГРАЦИЯ.

Тези и много други системи на компанията управляват широк спектър от данни - от структурирани таблични данни до сложни геометрични модели, като осигуряват интегрирано сътрудничество в процесите на проектиране, планиране и управление на строителството.

В примера за опростен диалог (фиг. 3.1-3) между специалистите по строителни проекти се обменят различни видове данни:

- ⦿ **Архитект:** "Като взех предвид желанието на клиента, добавих кът за сядане на покрива. Моля, разгледайте новия проект" (геометрични данни - модел).
- ⦿ **Строителен инженер:** "Проектът е получен. Изчислявам носещата способност на покрива за новата зона за отдих" структурирани и полуструктурни данни - таблици за изчисления).
- ⦿ **Ръководител на отдел "Покупки":** "Необходими са спецификации и количества материали за зоната за отдих, за да се организира покупката" (текстови и полуструктурни данни - списъци и спецификации).
- ⦿ **Инженер по здравословни и безопасни условия на труд:** "Получихте данни за новата зона. Оценявам рисковете и актуализирам плана за безопасност" полуструктурни данни - документи и планове).
- ⦿ **Специалист по BIM - моделиране:** "Извършване на промени в цялостния модел на проекта с цел адаптиране на работната документация" (геометрични данни и полуструктурни данни).
- ⦿ **Ръководител на проекта:** "Включвам новата зона за почивка в работния график. Актуализирам графиците и ресурсите в системата за управление на проекта" (структурни и полуструктурни данни - графици и планове).
- ⦿ **Специалист по поддръжка на съоръженията (FM):** "Подготвям данни за бъдещата поддръжка на зоната за отдих и ги въвеждам в системата за управление на собствеността" (структурни и полуструктурни данни - инструкции и планове за поддръжка).



Фиг. 3.1-3 Комуникацията между специалистите се осъществява както на ниво текст, така и на ниво данни.

Всеки специалист работи с различни видове данни, за да осигури ефективно сътрудничество в

екипа и успех на проекта. Разбирането на разликите между структурирани, полуструктурни и неструктурни данни ви позволява да разпознаете уникалната роля, която всеки тип играе в цифровите бизнес процеси. Важно е не само да се знае, че има различни форми на данни, но и да се разбере как, къде и защо се използват те.

Неотдавна идеята за комбиниране на толкова разнообразни данни изглеждаше амбициозна, но трудно осъществима. Днес тя вече е част от ежедневната практика. Интегрирането на данни с различни схеми и структури се е превърнало в неразделна част от архитектурата на съвременните информационни системи.

В следващите глави ще разгледаме подробно ключовите стандарти и подходи, които дават възможност за обединяване на структурирани, полуструктурни и неструктурни данни в един цялостен изглед. Особено внимание ще бъде обърнато на структурираните данни и релационните бази данни като основни механизми за съхраняване, обработка и анализ на информацията в строителната индустрия.

## Структурирани данни

В строителния бранш информацията идва от много източници - чертежи, спецификации, графики и доклади. За да се управлява ефективно този поток от информация, тя трябва да бъде структурирана. Структурираните данни ви позволяват да организирате информацията в удобна, четлива и достъпна форма.

Според 5-ия годишен доклад за технологиите в строителството на JB Knowledge [17] 67% от специалистите по управление на строителни проекти следят и оценяват изпълнението на работата ръчно или с помощта на електронни таблици.

Някои от най-разпространените формати за структурирани данни са XLSX и CSV. Те се използват широко за съхраняване, обработка и анализ на информация в електронни таблици. В такива електронни таблици данните се представят под формата на редове и колони, което ги прави лесни за четене, редактиране и анализиране.

**XLSX, формат**, създаден от Microsoft, се основава на използването на XML -структури и се архивира с помощта на алгоритъма ZIP. Основните характеристики на формата:

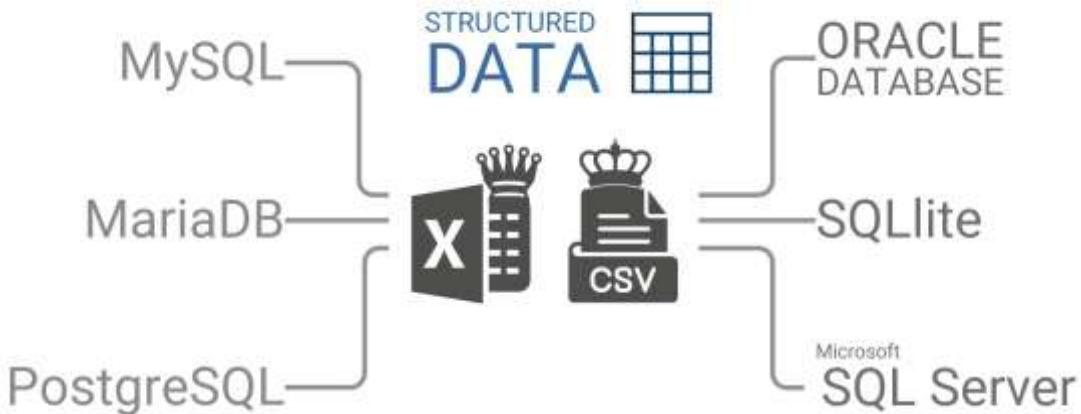
- Поддръжка на сложни формули, диаграми и макроси.
- Възможност за съхраняване на данни в различни листове, както и за форматиране на информацията.
- Оптимизиран за Microsoft Excel, но е съвместим и с други офис пакети.

**CSV формат** е обикновен текстов файл, в който стойностите са разделени със запетай, точки и подточки или други разделителни знаци. Основни предимства:

- Универсална съвместимост с различни програми и операционни системи.
- Лесно импортиране/експортиране към бази данни и аналитични системи.
- Лесна обработка дори в текстови редактори.

CSV обаче не поддържа формули и форматиране, така че основното му приложение е обмен на данни между системи и масово актуализиране на информация. Благодарение на своята универсалност и платформена независимост CSV се превърна в популярен инструмент за трансфер на данни в хетерогенни ИТ среди.

Двата формата XLSX и CSV служат за връзка между различни системи, работещи със структурирани данни (фигура 3.1-4). Те са особено полезни при задачи, при които са важни четливостта, ръчното редактиране и основната съвместимост.



**Фиг. 3.1-4** Форматите XLSX и CSV са връзката между различни системи, които работят със структурирани данни.

Независимостта от платформата прави CSV най-популярния формат за прехвърляне на данни в хетерогенни ИТ среди и системи.

Въпреки това XLSX и CSV не са предназначени за високопроизводителни изчисления или дългосрочно съхранение на големи количества данни. За такива цели се използват по-съвременни структурирани формати като Apache Parquet, Apache ORC, Feather, HDF5. Тези формати ще бъдат разгледани по-подробно в главата "Съхраняване на големи обеми данни: анализ на популярните формати и тяхната ефективност" в част 9 на тази книга.

На практика Excel с формат XLSX се използва по-често за малки задачи и автоматизиране на рутинни процеси. По-сложните сценарии изискват използването на системи за управление на данни, като ERP, PMIS, CAFM, CPM, SCM и други (фиг. 3.2-1). В тези системи се съхраняват структурирани данни, на които се основава организацията и управлението на информационните потоци в компанията.

Съвременните информационни системи за управление на данни, използвани в строителната индустрия, разчитат на структурирани данни, организирани под формата на таблици. За надеждно, мащабирано и цялостно управление на големи обеми информация разработчиците на приложения

и системи се обръщат към релационните системи за управление на бази данни (RDBMS).

## Релационни бази данни RDBMS и език за заявки SQL

За ефективно съхранение, обработка и анализ на данни **релационните бази данни (СУБД)** са системи за съхранение на данни, които организират информацията в таблици с определени връзки между тях.

Данните, организирани в бази данни (СУБД), не са просто цифрова информация; те са основа за транзакции и взаимодействия между различни системи.

Ето няколко от най-разпространените релационни системи за управление на бази данни (RDBMS) (Фигура 3.1-5):

- **MySQL** (с отворен код) е една от най-популярните СУБД, която е част от стека LAMP (Linux, Apache, MySQL, PHP /Perl/Python). Тя е широко използвана при разработването на уеб сайтове поради своята простота и висока производителност.
- **PostgreSQL** (с отворен код) е мощна обектно-релационна система, известна със своята надеждност и разширени функции. Тя е подходяща за сложни корпоративни решения.
- **Microsoft SQL Server** е търговска система на Microsoft, която се използва широко в корпоративни среди поради интеграцията си с други продукти на компанията и високото ниво на сигурност.
- **Oracle Database** е една от най-мощните и надеждни СУБД, използвани в големи предприятия и критични приложения.
- **IBM DB2** - предназначена за големи корпорации, осигуряваща висока производителност и устойчивост на грешки.
- **SQLite** (с отворен код) е олекотена вградена база данни, идеална за мобилни приложения и самостоятелни системи, като например софтуер за CAD проектиране (BIM).

Популярните системи за управление на бази данни в строителния бизнес - MySQL, PostgreSQL, Microsoft SQL Server, Oracle® Database, IBM® DB2 и SQLite - работят със структурирани данни.

Всички тези СУБД са мощни и гъвкави решения за управление на широк спектър от бизнес процеси и приложения - от малки уеб сайтове до мащабни корпоративни системи (фиг. 3.2-1).

Според Statista [48] през 2022 г. релационните системи за управление на бази данни (RDBMS) ще представляват около 72% от всички използвани СУБД.

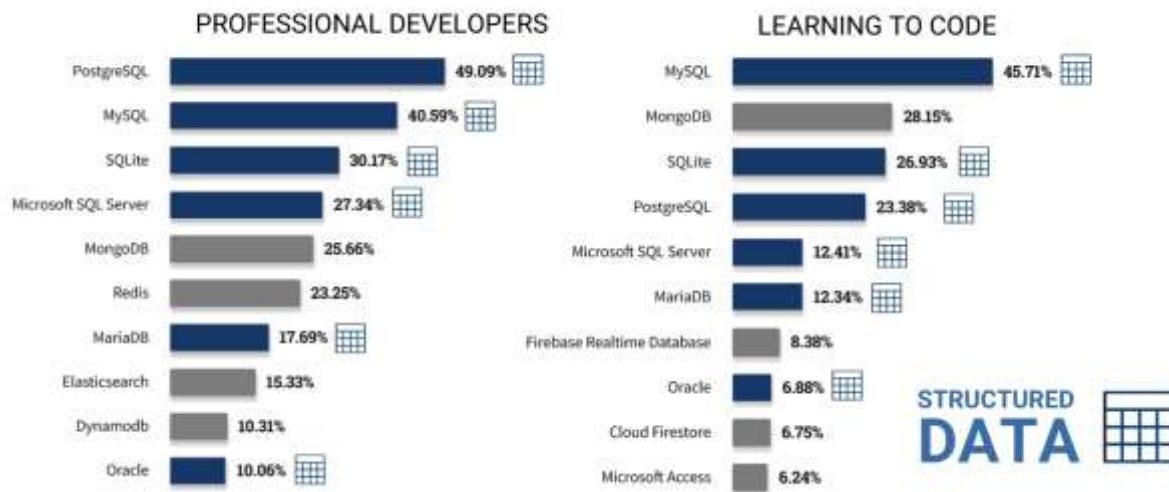
	Rank			DBMS	Database Model	Open Source vs Commercial
	Mar2025	Feb2025	Mar2024			
	1.	1.	1.	Oracle®	Relational, Multi-model	Commercial
	2.	2.	2.	MySQL	Relational, Multi-model	Open Source
	3.	3.	3.	Microsoft® SQL Server	Relational, Multi-model	Commercial
	4.	4.	4.	PostgreSQL	Relational, Multi-model	Open Source
	5.	5.	5.	MongoDB	Document, Multi-model	Open Source
	6.	7.	9.	Snowflake®	Relational	Commercial
	7.	6.	6.	Redis®	Key-value, Multi-model	Open Source
	8.	8.	7.	Elasticsearch®	Multi-model	Open Source
	9.	9.	8.	IBM Db2	Relational, Multi-model	Commercial
	10.	10.	10.	SQLite	Relational	Open Source
	11.	11.	12.	Apache Cassandra®	Multi-model	Open Source
	12.	12.	11.	Microsoft Access®	Relational	Open Source
	13.	13.	17.	Databricks®	Multi-model	Commercial
	14.	14.	13.	MariaDB	Relational, Multi-model	Open Source
	15.	15.	14.	Splunk	Search engine	Commercial
	16.	16.	16.	Amazon DynamoDB	Multi-model	Commercial
	17.	17.	15.	Microsoft Azure SQL	Relational, Multi-model	Commercial

Фиг. 3.1-5 Популярност на използването на структурирани бази данни (отбелязани в синьо) класацията на СУБД (въз основа на [49]).

Инсталирането на бази данни с отворен код е доста лесно - дори без задълбочени технически познания. Системите с отворен код, като PostgreSQL, MySQL или SQLite, са достъпни бесплатно и работят на повечето операционни системи: Windows, MacOS и Linux. Необходимо е само да отидете на официалния уебсайт на проекта, да изтеглите инсталатора и да следвате инструкциите. В повечето случаи инсталацията отнема не повече от 10-15 минути. В четвъртата част на книгата ще моделираме и създадем една такава база данни (фиг. 4.3-8).

Ако компанията ви използва облачни услуги (например Amazon Web Services, Google Cloud или Microsoft Azure), можете да разгърнете базата данни с няколко кликвания - платформата ще ви предложи готови шаблони за инсталација. Благодарение на отвореността на кода, такива бази данни са лесни за персонализиране за вашите задачи, а огромната общност от потребители винаги ще ви помогне да намерите решение на всеки проблем.

СУБД остават в основата на множество бизнес приложения и платформи за анализ (фигура 3.1-6), които позволяват на компаниите ефективно да съхраняват, обработват и анализират данни - и следователно да вземат информирани и навременни решения.



Фиг. 3.1-6 Проучване сред разработчици в StackOverFlow (най-големия ИТ форум) за това кои бази данни са използвали миналата година и кои искат да използват през следващата година (СУБД са подчертани в синьо) (въз основа на [50]).

СУБД осигуряват надеждност, съгласуваност на данните, поддръжка на транзакции и използват мощен език за заявки - SQL (Structured Query Language), който често се използва в анализите и позволява лесно да получавате, променяте и анализирате информацията, съхранявана в базите данни. SQL е основният инструмент за работа с данни в релационни системи.

## SQL - заявки в базите данни и нови тенденции

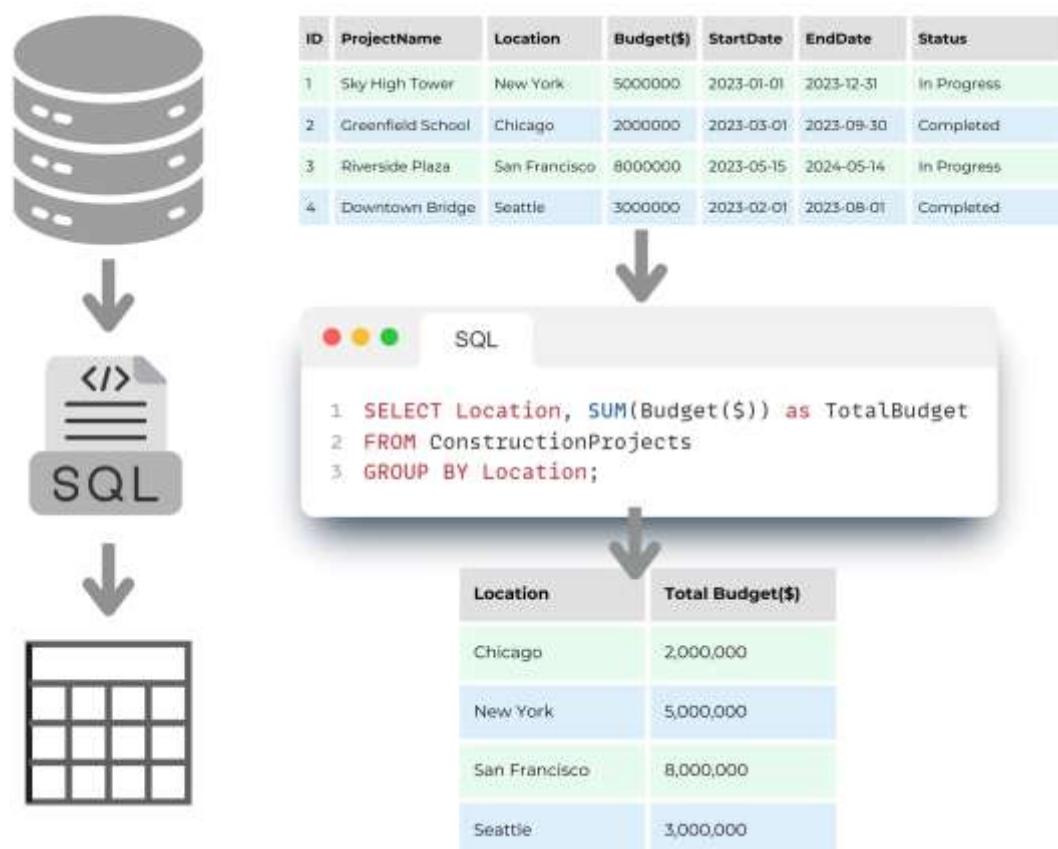
Основното предимство на езика SQL, често използван в релационни бази данни, пред други видове управление на информацията (например с помощта на класическите електронни таблици Excel) е поддържането на много големи обеми бази данни при висока скорост на обработка на заявките.

Структурираният език за заявки (SQL) е специализиран език за програмиране, предназначен за съхраняване, обработка и анализ на информация в релационни бази данни. SQL се използва за създаване, управление и достъп до данни, като ви позволява ефективно да намирате, филтрирате, комбинирате и обобщавате информация. Той служи като ключов инструмент за достъп до данни, осигурявайки удобен и формализиран начин за взаимодействие с информационните хранилища.

Еволюцията на системите SEQUEL-SQL преминава през значими продукти и компании, като Oracle, IBM DB2, Microsoft SQL Server, SAP, PostgreSQL и MySQL, и завършва с появата на SQLite и MariaDB [51]. SQL предоставя възможности за работа с електронни таблици, каквито не се срещат в Excel, което прави манипулирането на данни по-скалируемо, сигурно и лесно за автоматизиране:

- **Създаване и управление на структури от данни (DDL):** В SQL можете да създавате, променяте и изтривате таблици в базата данни, да установявате връзки между тях и да определяте структури за съхранение на данни. Excel, от друга страна, работи с фиксирани листове и клетки, без ясно дефинирани връзки между листове и набори от данни.

- **Манипулиране на данни (DML):** SQL ви позволява масово да добавяте, променяте, изтривате и извличате данни с висока скорост, като изпълнявате сложни заявки с филтриране, сортиране и обединяване на таблици (фигура 3.1-7). В Excel обработката на големи количества информация изисква ръчни действия или специални макроси, което забавя процеса и увеличава вероятността от грешки.
- **Контрол на достъпа (DCL):** SQL ви позволява да диференцирате правата за достъп до данни за различни потребители, като ограничавате възможността за редактиране или преглед на информация. В Excel, от друга страна, достъпът е или споделен (при прехвърляне на файл), или изисква сложни настройки със споделяне на права чрез облачни услуги.



Фиг. 3.1-7 Пример за DML в SQL: бърза обработка, групиране и агрегиране с няколко реда код за автоматична обработка на данни.

Excel улеснява работата с данни с визуалната си и интуитивна структура. Въпреки това, с увеличаването на обема на данните, производителността на Excel намалява. Excel също така се сблъсква с ограничения за обема на данните, които може да съхранява - максимум един милион реда - и производителността се влошава много преди достигането на този лимит. Така че, докато Excel изглежда за предпочитане за визуализиране и манипулиране на малки количества данни, SQL е по-подходящ за работа с големи масиви от данни.

Следващият етап в развитието на структурираните данни е появата на колонарните бази данни

(Columnar Databases), които са алтернатива на традиционните релационни бази данни, особено когато става въпрос за значително по-големи обеми данни и аналитични изчисления. За разлика от редовите бази данни, при които данните се съхраняват ред по ред, при колонните бази данни информацията се записва по колони. В сравнение с класическите бази данни това позволява:

- Намалете пространството за съхранение чрез ефективно компресиране на еднотипни данни в колони.
- Ускоряване на аналитичните заявки, тъй като се четат само необходимите колони, а не цялата таблица.
- Оптимизиране на големите данни и складирането на данни, напр. Data Lakehouse Architecture.

Ще говорим повече за колонарните бази данни, Pandas DataFrame, Apache Parquet, HDF5, както и за създаването на базирани на тях Big Data -хранилища за целите на анализа и обработката на данни в следващите глави на тази книга - "DataFrame: универсален табличен формат за данни" и "Формати за съхранение на данни и работа с Apache Parquet: DWH -хранилища за данни и архитектура на Data Lakehouse".

## Неструктурирани данни

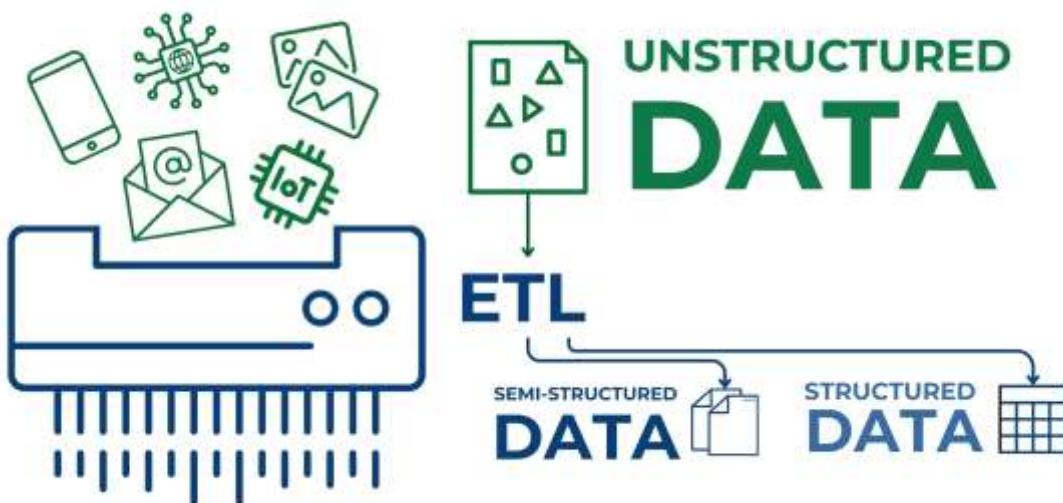
Въпреки че повечето данни, използвани в приложенията и информационните системи, са в структуриран вид, по-голямата част от информацията, генерирана в строителството, е под формата на неструктурирани данни - изображения, видеоклипове, текстови документи, аудиозаписи и други форми на съдържание. Това важи особено за етапите на строителство, експлоатация и технически надзор, където преобладава визуалната и текстовата информация.

**Неструктурираните данни** са информация, която няма предварително определен модел или структура, не е организирана в традиционни редове и колони, както в базите данни или таблиците.

Най-общо казано, неструктурираните данни могат да бъдат класифицирани в две категории:

- Неструктурирани данни, генериирани от човека, които включват различни видове съдържание, генерирано от човека: текстови документи, имейли, изображения, видеоклипове и т.н.
- Неструктурираните данни, генериирани от машини, се създават от устройства и сензори: те включват файлове с регистри, GPS данни, резултати от интернет на нещата (IoT) и друга телеметрична информация от строителна площадка, например.

За разлика от структурираните данни, които са удобно организирани в таблици и бази данни, неструктурираните данни изискват допълнителни стъпки за обработка преди интегрирането им в информационните системи (фиг. 3.1-8). Използването на технологии за автоматизирано събиране, анализ и преобразуване на такива данни открива нови възможности за подобряване на ефективността на строителството, намаляване на грешките и минимизиране на влиянието на човешкия фактор.



**Фигура 3.1-8** Обработката на неструктурирани данни започва с тяхното трансформиране в полуструктурни и структурирани данни.

Неструктурираните данни съставляват до 80% от цялата информация [52], с която се сблъскват специалистите в компаниите, затова в следващите глави на книгата ще разгледаме подробно техните видове и обработка с примери.

За улеснение на дискусията текстовите данни са категоризирани отделно. Въпреки че те са вид по-скоро неструктурни данни, тяхното значение и разпространение в строителната индустрия изискват специално внимание.

### Текстови данни: между неструктурен хаос и структуриран хаос

Текстовите данни в строителния бранш обхващат широк спектър от формати и видове информация - от документи на хартия до неформални методи на комуникация, като писма, разговори, работна кореспонденция и устни срещи на строителната площадка. Всички тези текстови данни носят важна информация за управлението на строителните проекти - от подробности за проектни решения и промени в плановете до обсъждане на въпроси, свързани с безопасността, и преговори с изпълнители и клиенти (фигура 3.1-9).



**Фиг. 3.1-9** Текстови данни, един от най-популярните видове информация, използвани при комуникацията между участниците в проекта.

Текстовата информация може да бъде както формализирана, така и неструктуррирана. Формализираните данни включват документи в Word (.doc,.docx), PDF, както и текстови файлове с протоколи от срещи (.txt). Неформализираните данни включват кореспонденция по месингджър и електронна поща, стенограми от срещи (Teams, Zoom, Google Meet) и аудиозаписи на дискусии (.mp3,.wav), които изискват преобразуване в текст.

Но докато писмените документи, като официални искания, договорни условия и имейли, обикновено вече имат определена структура, устните съобщения и работната кореспонденция често остават неструктурирани, което затруднява анализирането и интегрирането им в системите за управление на проекти.

Ключът към ефективното управление на текстови данни е да ги превърнете в структуриран формат. Това позволява обработената информация да бъде автоматично интегрирана в съществуващите системи, които вече работят със структурирани данни.



Фиг. 3.1-10 Конвертиране на текстово съдържание в структурирани данни.

За да се използва ефективно текстовата информация, тя трябва да се преобразува автоматично в структурирана форма (фигура 3.1-10). Този процес обикновено включва няколко стъпки:

- **Разпознаване на текст (OCR)** - преобразуване на изображения на документи и чертежи в машинночетим формат.
- **Анализ на текста (NLP)** - автоматично идентифициране на ключови параметри (дати, суми и цифри, свързани с проекта).
- **Класификация на данни** - категоризиране на информация (финанси, логистика, управление на риска).

След разпознаване и класифициране вече структурираните данни могат да се интегрират в бази данни и да се използват в автоматизирани системи за отчитане и управление.

## Полуструктурирани и слабо структурирани данни

Полуструктурирани данни съдържа известно ниво на организация, но няма строга схема или структура. Въпреки че подобна информация включва структурирани елементи (напр. дати, имена на служителите и списъци на изпълнените задачи), форматът на представяне може да варира значително

в различните проекти или дори между отделните служители. Примери за такива данни са дневниците за отчитане на работното време, докладите за напредъка и графиците, които могат да бъдат представени в различни формати.

Полуструктурираните данни се анализират по-лесно от неструктурните данни, но изискват допълнителна обработка за интегриране в стандартизиранi системи за управление на проекти.

Работата с полуструктурни данни, характеризиращи се с постоянно променяща се структура, представлява сериозно предизвикателство. Това е така, защото променливостта на структурата на данните изисква отделни индивидуални подходи за обработка и анализ на всеки източник на полуструктурни данни.

Но докато работата с неструктурни данни изисква много усилия, обработката на полуструктурни данни може да се извърши със сравнително прости методи и инструменти.

**Слабо структурирани** данни е по-общ термин, който описва данни с минимална или непълна структура. Най-често това са текстови документи, чатове, имейли, в които се откриват някои метаданни (напр. дата, изпращач), но по-голямата част от информациите е представена хаотично.

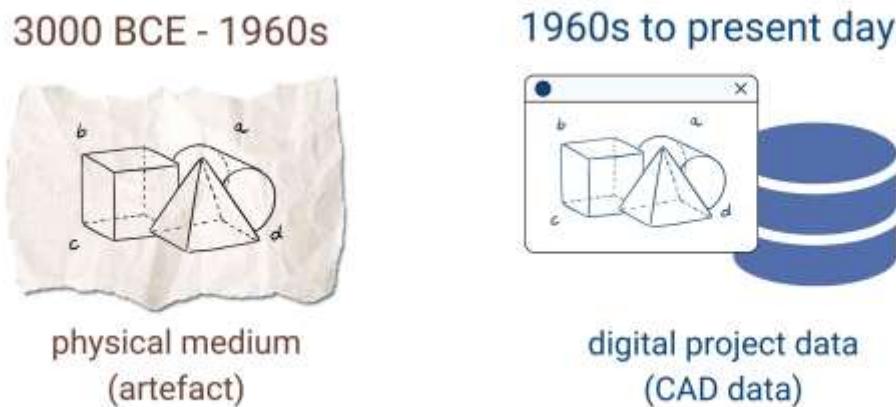
В строителството свободно структурираните данни се срещат в различни процеси. Те могат да включват например:

- Оценки и оферти - таблици с данни за материали, количества и разходи, но без единен формат.
- Чертежи и инженерни диаграми - файлове в PDF или DWG, съдържащи текстови анотации и метаданни, но без строго фиксирана структура.
- Работни графици - данни от MS Project, Primavera P6 или други системи, които може да имат различна структура на експортиране.
- CAD (BIM-модели) - съдържат елементи на конструкцията, но представянето на данните зависи от софтуера и стандарта на проекта.

Геометричните данни, създадени от CAD системите, могат да бъдат категоризирани по същия начин като полуструктурните данни. Въпреки това ще класифицираме геометричните CAD (BIM) данни като отделен тип данни, тъй като те, подобно на текстовите данни, често могат да бъдат третирани като отделен тип данни в процесите на компанията.

## Геометрични данни и тяхното приложение

Докато метаданните за елементите на проекта почти винаги се съхраняват под формата на таблици, структурирани или слабо структурирани формати, геометричните данни за елементите на проекта в повечето случаи се създават с помощта на специални CAD инструменти (фиг. 3.1-11), които позволяват детайлно визуализиране на елементите на проекта като набор от линии (2D) или геометрични тела (3D).

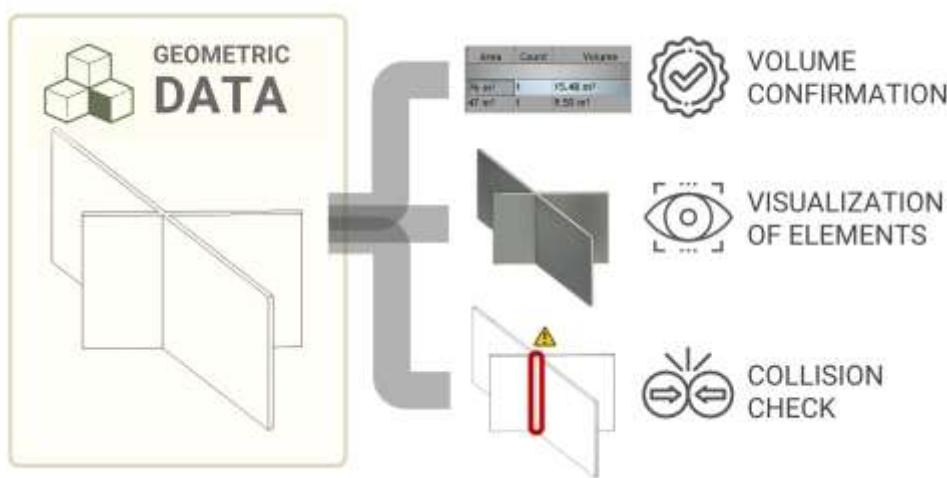


**Фиг. 3.1-11 CAD инструментите са помогнали за пренасянето на геометричната информация от физически носител във вид на база данни.**

При работа с геометрични данни в строителството и архитектурата могат да се определят три основни приложения на геометричните данни (фигура 3.1-12):

- **Потвърждаване на обемите:** геометричните данни, генериирани в рамките на CAD програмите (BIM) с помощта на специални геометрични ядра, са необходими за автоматично и точно определяне на обемите и размерите на елементите на проекта. Тези данни включват автоматично изчислени площи, обеми, дължини и други важни атрибути, необходими за планирането, бюджетирането и поръчването на ресурси и материали
- **Визуализация на проекта:** в случай на промени в проекта, визуализацията на елементите позволява автоматично генериране на актуализирани чертежи в различни равнини. Визуализацията на проекта в началните етапи помага да се ускори разбирането между всички участници, за да се спестят време и ресурси по време на строителния процес.
- **Проверка на сблъсъци:** В сложни строителни и инженерни проекти, където взаимодействието на множество категории елементи (напр. тръби и стени) без "геометрични конфликти" е от решаващо значение, проверката на сблъсъци играе ключова роля. Използването на софтуера за откриване на сблъсъци ви позволява да идентифицирате проактивно потенциални геометрични конфликти между елементите на проекта, предотвратявайки скъпоструващи грешки по време на строителния процес.

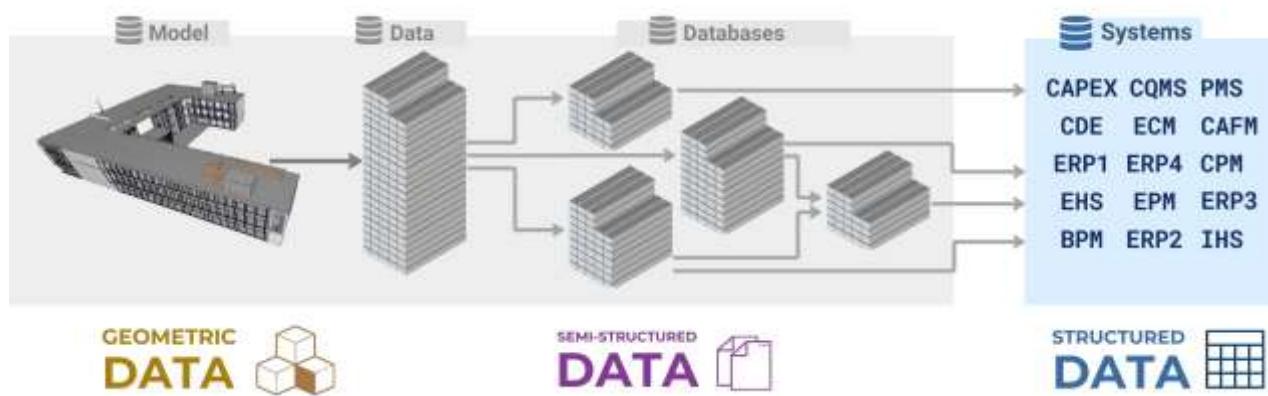
От самото начало на инженерните проектантски бюра, от времето на изграждането на първите сложни конструкции, строителните инженери са предоставяли геометрична информация под формата на чертежи, линии и плоски геометрични елементи (на папирус, хартия "A0" или във формат DWG, PDF, PLT), въз основа на които майсторите и оценителите (фиг. 3.1-11), през последните хилядолетия с помощта на линийки и транспортири са събирали атрибутивни обеми или брой елементи и групи елементи (3.1-11), за последните хилядолетия, с помощта на линийки и транспорти, са събирали атрибутивни обеми или количества на елементи и групи елементи.



**Фиг. 3.1-12** Геометрията е основата за получаване на обемните параметри на елементите, които след това се използват за изчисляване на разходите и сроковете на проекта.

Днес тази ръчна и отнемаша време задача е решена напълно автоматизирано благодарение на появата на обемното моделиране в съвременните CAD инструменти (BIM), което позволява автоматично, с помощта на специално геометрично ядро, да се получат обемните атрибути на всеки елемент, без да е необходимо обемните параметри да се изчисляват ръчно.

Съвременните CAD инструменти позволяват също така да класифицирате и категоризирате елементите на проекта, така че да можете да качвате таблици със спецификации от базата данни на проекта за използване в различни системи, като например за оценка на разходите, планиране или изчисляване на CO<sub>2</sub> (фиг. 3.1-13). Получаването на спецификации, таблици за QTO и количества, както и практически примери ще разгледаме в главата "Получаване на количества и количествено изчисляване".



**Фиг. 3.1-13** CAD инструментите (BIM) съхраняват данни в бази данни, които са проектирани да се интегрират и да взаимодействват с други системи.

Поради затворения характер на базите данни и форматите, използвани в CAD средата, геометричните данни, създадени в CAD решенията, всъщност са се превърнали в отделен вид информация.

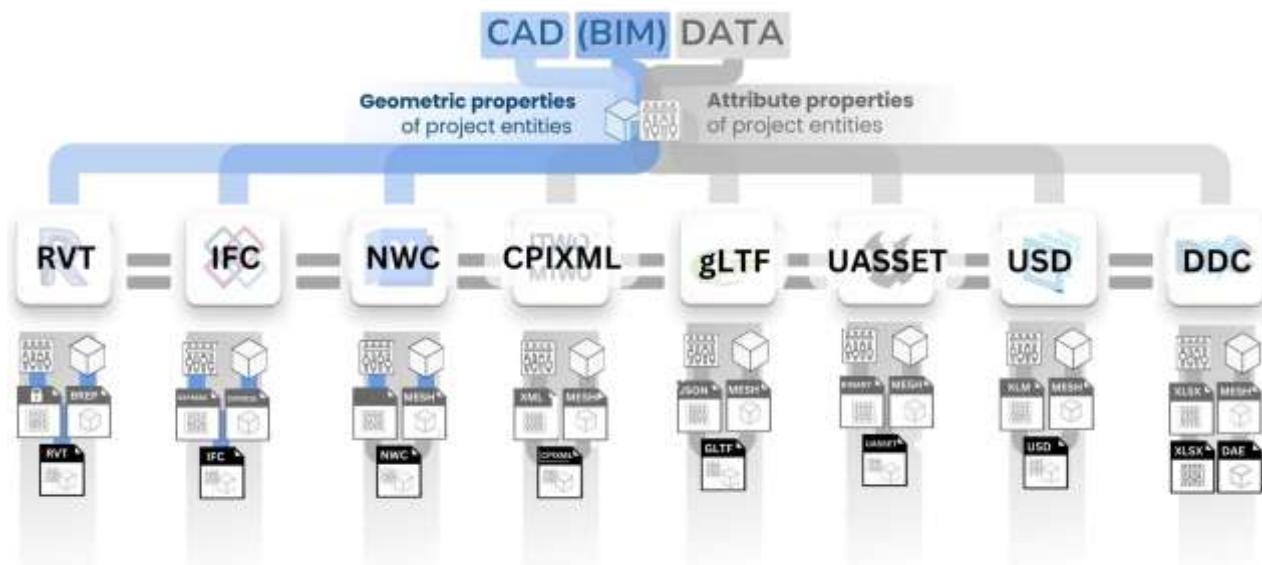
Тя съчетава както геометрията на елементите, така и метаданната (структурни или полу-структурни) информация (струтурирана или полу-структурна), затворена в специализирани файлове и формати.

## CAD данни: от проектирането до съхранението на данни

Съвременните CAD и BIM системи съхраняват данни в свои собствени, често патентовани формати: DWG, DXF, RVT, DGN, PLN и други. Тези формати поддържат както 2D, така и 3D представяне на обектите, като запазват не само геометрията, но и атрибути, свързани с обектите. Ето най-разпространените от тях:

- **DWG** е двоичен файлов формат, използван за съхраняване на двуизмерни (и по-рядко триизмерни) проектни данни и метаданни.
- **DXF** е текстов формат за обмен на двуизмерни и триизмерни -чертежи между CAD -системи. Той съдържа данни за геометрия, слоеве и атрибути, поддържа както ASCII, така и двоично представяне.
- **RVT** е двоичен формат за съхраняване на CAD модели, включващи 3D -геометрия, атрибути на елементите, връзки и параметри на дизайна.
- **IFC** е отворен текстов формат за обмен на строителни данни между CAD (BIM) системи. Той включва геометрия, свойства на обектите и информация за техните взаимоотношения.

Освен тях се използват и други формати: PLN, DB1, SVF, NWC, CPIXML, BLEND, BX3, USD, XLSX, DAE. Въпреки че се различават по предназначение и степен на отвореност (фигура 3.1-14), всички те могат да представлят един и същ информационен модел на проекта в различни форми. В сложните проекти тези формати често се използват паралелно - от изготвянето до координирането на моделите на проекта.



Фигура 3.1-14 Популярните CAD формати за съхранение описват геометрията чрез BREP или MESH параметри, допълнени от атрибутни данни.

Всички горепосочени формати позволяват да се съхраняват данни за всеки елемент от строителния проект и всички горепосочени формати съдържат два основни типа данни:

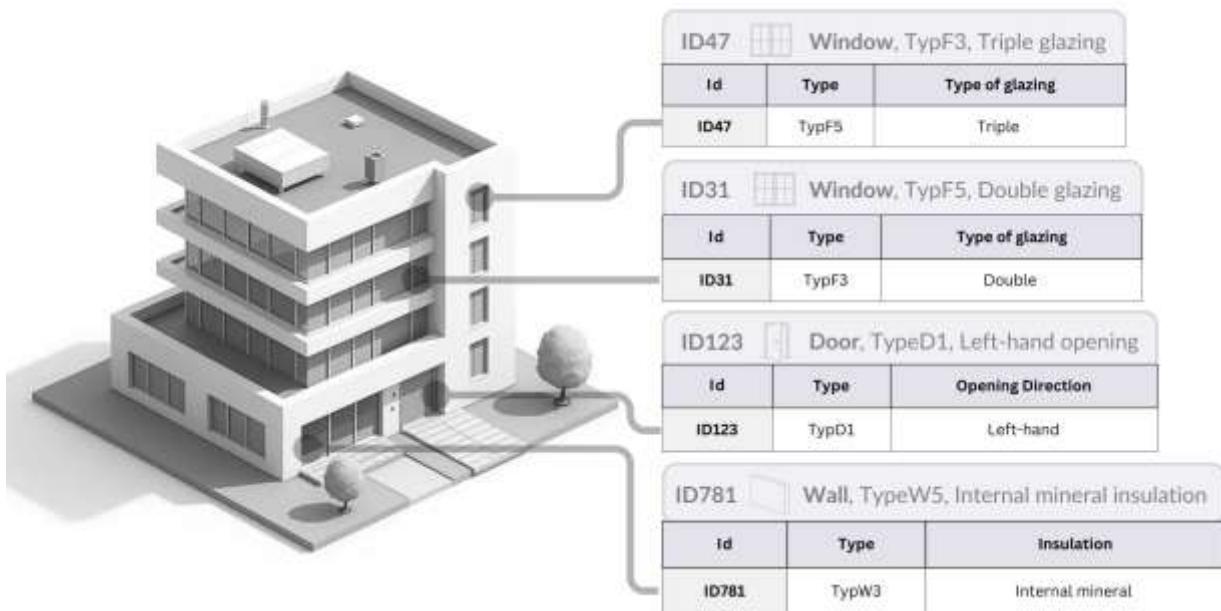
- **Геометрични параметри** - описват формата, местоположението и размерите на даден обект. Геометрията и нейното използване ще бъдат разгледани подробно в шестата част на книгата, посветена на CAD (BIM) решенията;
- **Свойства на атрибутите** - съдържат различна информация: материали, типове елементи, технически характеристики, уникални идентификатори и други свойства, които елементите на проекта могат да притежават.

Данните за атрибутите са от особено значение в съвременните проекти, тъй като те определят експлоатационните характеристики на обектите, позволяват инженерни изчисления и изчисляване на разходите и осигуряват цялостно взаимодействие между участниците в проектирането, строителството и експлоатацията. Например:

- За прозорци и врати: вид на конструкцията, вид на остъкляването, посока на отваряне (фиг. 3.2-1).
- За стените се записва информация за материалите, топлоизолацията и акустичните характеристики.
- За инженерните системи се съхраняват параметрите на тръбопроводите, каналите, кабелните трасета и техните връзки.

Тези параметри могат да се съхраняват както в самите CAD-(BIM)-файлове, така и във външни бази данни - в резултат на експорт, преобразуване или директен достъп до вътрешните CAD структури чрез инструменти за обратно инженерство. Този подход улеснява интегрирането на информациите за проектирането с други корпоративни системи и платформи

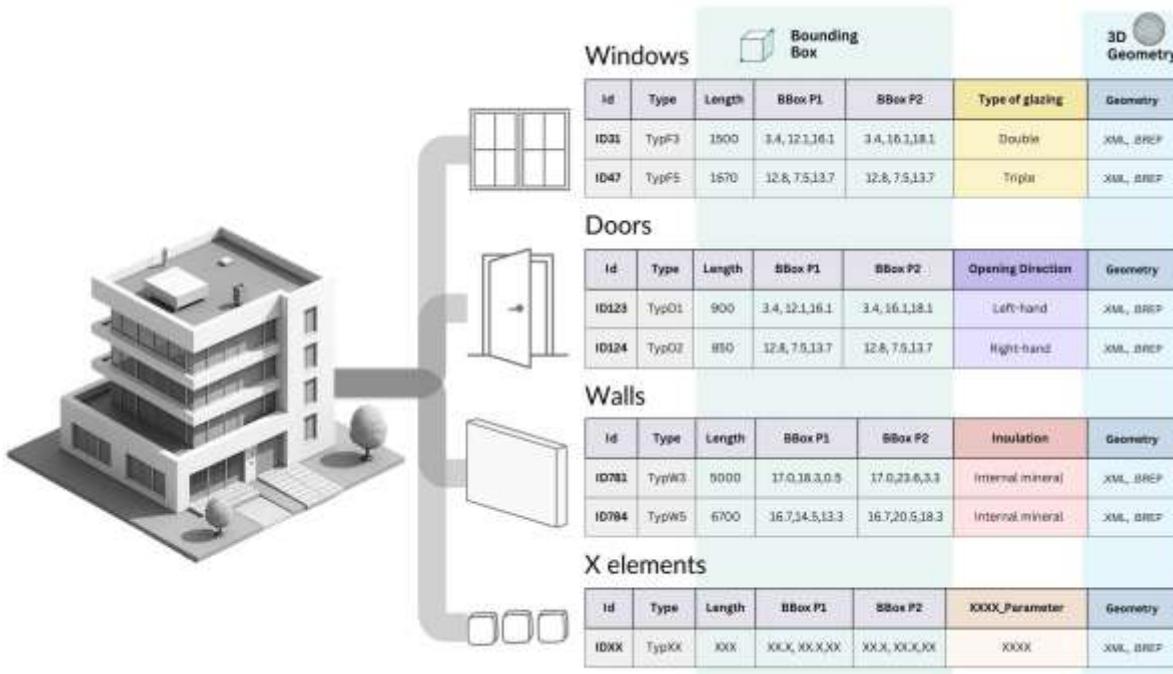
**Обратният инженеринг** в контекста на CAD (BIM) е процесът на извлечане и анализиране на вътрешната структура на цифров модел с цел пресъздаване на неговата логика, структура на данните и зависимости без достъп до оригиналните алгоритми или документация.



Фиг. 3.1-15 Елементът на проекта, освен че описва параметрична или полигонална геометрия, съдържа информация за параметрите и свойствата на елементите.

В резултат на това около всеки елемент се формира уникален набор от параметри и свойства, включващ както уникални характеристики на всеки обект (напр. идентификатор и размери), така и общи атрибути за групи елементи. Това позволява не само да се анализират отделните елементи-обекти на проекта, но и да се обединяват в логически групи, които след това могат да се използват от други специалисти за техните задачи и изчисления в системи и бази данни.

Същност е конкретен или абстрактен обект от реалния свят, който може да бъде едно-значно идентифициран, описан и представен под формата на данни.



Фиг. 3.1-16 Всеки елемент на проекта съдържа атрибути, които се въвеждат от проектанта или се изчисляват в програмата CAD.

През последните десетилетия строителната индустрия разработи много нови CAD (BIM) формати, които опростяват създаването, съхранението и предаването на данни. Тези формати могат да бъдат затворени или отворени, таблични, параметрични или графични. Тяхното разнообразие и фрагментарност обаче значително усложняват управлението на данните на всички етапи от жизнения цикъл на проекта. Сравнителна таблица на основните формати, използвани за обмен на информация в строителството, е представена на фигура 3.1-17 (пълната версия е достъпна чрез QR код).

За да се решат проблемите с оперативната съвместимост и достъпа до CAD данни, се включват BIM мениджъри (BIM) и координатори, чиято задача е да контролират износа, да проверяват качеството на данните и да интегрират части от CAD (BIM) данни в други системи.

Поради затворения характер и сложността на форматите обаче е трудно да се автоматизира този процес, което принуждава специалистите да извършват много операции ръчно, без да могат да изградят пълноценни поточни процеси за обработка на данни (pipeline).

The table is a comprehensive comparison chart for construction CAD (BIM) data storage. It lists numerous formats and their characteristics across several categories. Key columns include:

- Geometric properties of project entities:** Includes icons for points, lines, surfaces, and solids.
- Attribute properties of project entities:** Includes icons for text, numbers, dates, and other data types.
- Comparative formats from construction projects:** Lists various formats such as DWG, DXF, IFC, and others.
- Storage:** Describes how each format is stored, often mentioning 'as-is' or 'parsed'.
- Performance:** Evaluates speed, memory usage, and compatibility.
- Integration:** Details how well each format integrates with other systems.
- Cost:** Provides a rough estimate of cost.
- Complexity:** Rates the complexity of each format.
- Standardization:** Indicates if the format is widely adopted or standardized.
- Future:** Predicts the future of each format.

Фиг. 3.1-17 Таблица за сравнение на основните формати на данни, в които се съхранява информация за елементите на проекта [53].

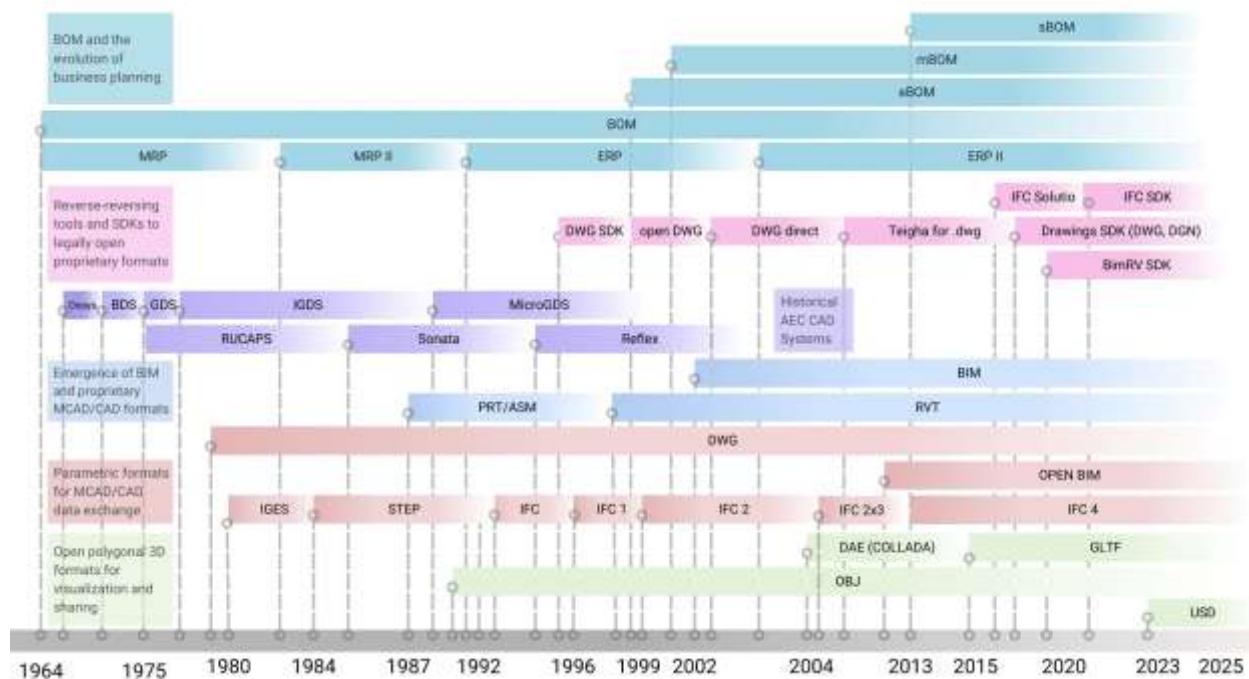
За да се разбере защо има толкова много различни формати на данни и защо повечето от тях са затворени, е важно да се навлезе в процесите, които протичат в CAD (BIM) програмите, които ще бъдат разгледани подробно в шестата част на книгата.

Допълнителен информационен слой, добавен към геометрията, е въведен от разработчиците на CAD системи под формата на концепцията BIM (Building Information Modeling) - маркетингов термин, активно популяризиран в строителната индустрия от 2002 г. насам [54].

### Появата на концепцията BIM (BOM) и използването на CAD в процесите

Концепцията за информационно моделиране на сградата (BIM), представена за първи път в бялата книга за BIM от 2002 г. [54], произлиза от маркетинговите инициативи на производителите на CAD софтуер. Тя възникна от маркетинговите инициативи на разработчиците на CAD софтуер и представляваше опит за адаптиране на вече утвърдени в машиностроенето принципи към нуждите на строителната индустрия.

Вдъхновението за BIM идва от концепцията за BOM (Bill of Materials) - спецификация на материалите, която се използва широко в промишлеността от края на 80-те години на миналия век. В машиностроенето BOM позволи свързването на данните от CAD системите с PDM (Product Data Management), PLM (Product Lifecycle Management) и ERP системите, осигурявайки цялостно управление на инженерната информация през целия жизнен цикъл на продукта (фиг. 3.1-8).



Фиг. 3.1-18 Еволюция на спецификациите (BOM), информационното моделиране (BIM) и цифровите формати в инженерната строителна индустрия.

Съвременното развитие на концепцията за ВОМ доведе до появата на разширена рамка - XBOM (Extended BOM), която включва не само състава на продукта, но и сценарии за поведение, оперативни изисквания, параметри за устойчивост и данни за прогнозен анализ. XBOM по същество изпълнява същата роля като BIM в строителството: и двата подхода се стремят да превърнат цифровия модел в Единен източник на истина за всички участници в проекта през целия му жизнен цикъл.

Ключов етап в появата на ВОМ в строителството е въвеждането на първия параметричен CAD (MCAD), специално адаптиран за строителната индустрия през 2002 г. Тя е разработена от екипа, който преди това е създал Pro-E® - революционна MCAD система за машиностроение, появила се в края на 80-те години и превърнала се в индустрислен стандарт [55].

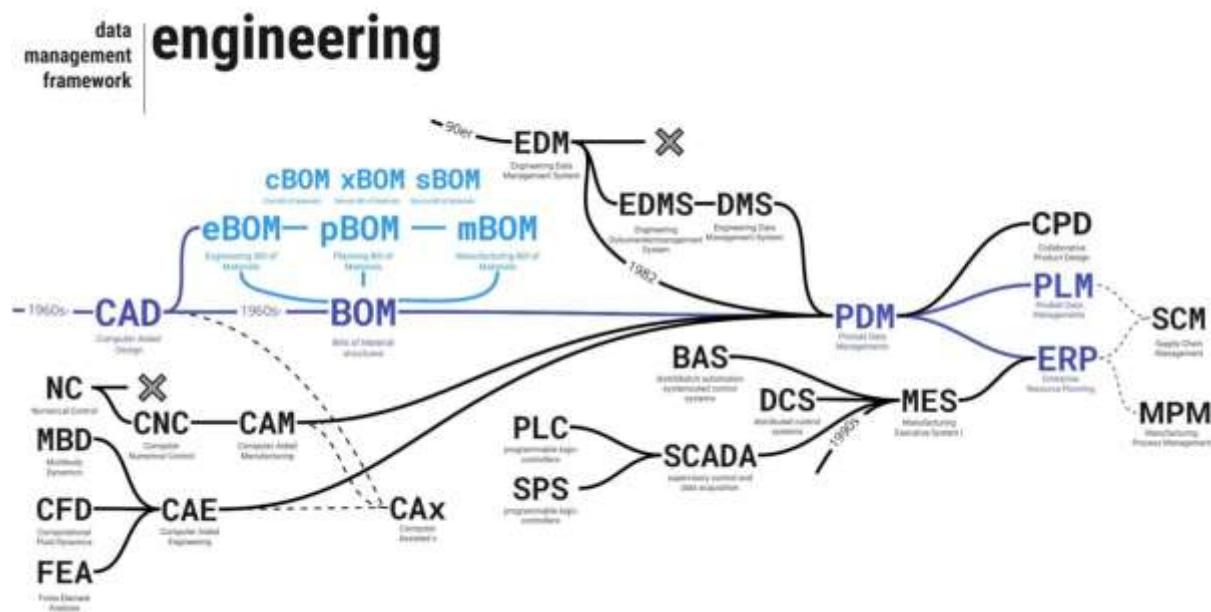
Още в края на 80-те години на миналия век целта беше да се премахнат ограниченията [56] на съществуващите тогава CAD-програми. Основната цел беше да се намали трудът, необходим за извършване на промени в параметрите на конструктивните елементи, и да се даде възможност за актуализиране на модела въз основа на данни извън CAD-програмите чрез база данни [57]. Най-важната роля за това трябваше да играе параметризацията: автоматично извлечане на характеристики от базата данни и използването им за актуализиране на модела в CAD системите.

Pro-E и концепцията за елементарно параметрично моделиране с ВОМ, която стои в основата му, оказаха значително влияние върху развитието на пазара на CAD- и MCAD [58]. В продължение на 25 години този модел е бил в индустрията и много съвременни системи са станали негови концептуални наследници.

Целта е да се създаде система, която е достатъчно гъвкава, за да насърчи инженера лесно да разглежда различни проекти. А разходите за извършване на промени в проекта трябва да са възможно най-близки до нулата. Традиционният CAD / CAM софтуер нереалистично ограничава извършването на евтини промени само в самото начало на процеса на проектиране [59].

- Самуел Хайзенберг, основател на Parametric Technology Corporation®, разработчик на MCAD - продукта Pro-E и преподавател на създателя на CAD продукт, използващ RVT формат

В машиностроенето системите PDM, PLM, MRP и ERP се превърнаха в ключови платформи. Те играят централна роля в управлението на данните и процесите, като събират информация от CAx системите (CAD, CAM, CAE) и организират дейностите по проектиране въз основа на структурата на продукта (BOM: eBOM, pBOM, mBOM) (фиг. 3.1-18). Тази интеграция намалява грешките, избягва дублирането на данни и осигурява проследимост от край до край от проектирането до производството.



**Фигура 3.1-19 Исторически BOM възниква през 60-те години на миналия век като начин за структуриране на данните от CAx системите и предаването им на системите за управление.**

Закупуването от един от водещите доставчици на CAD решение, разработено от бившия екип на Pro-E и базирано на подхода BOM, беше отбелязано с почти незабавното публикуване на серията BIM Whitepaper (2002-2003 г.)[60][61]. Още от средата на 2000-те години концепцията BIM започна да се популяризира активно в строителната индустрия, което значително повиши интереса към па-

раметричния софтуер. Популярността нарасна толкова бързо, че строителното разклонение на машиностроенето Pro-E - параметричен CAD, промотиран от този производител - всъщност измести конкурентите в сегмента на архитектурното и конструктивното проектиране (фиг. 3.1-20). До началото на 2020-те години той де факто е затвърдил глобалното си господство на пазара на BIM (CAD) [62].



Фиг. 3.1-20 Популярност на заявките за търсене в Google (RVT спрямо IFC): параметричният CAD, създаден от бившия екип на Pro-E с поддръжка на BOM -BIM, е придобил популярност в почти повечето страни по света.

През последните 20 години съкращението BIM придоби множество тълкувания, чито многобройни значения се коренят в първоначалните маркетингови концепции, възникнали в началото на 2000-те години. Стандартът ISO 19650, който изигра важна роля за популяризирането на термина, всъщност осигури статута на BIM като "научно обоснован" подход към управлението на информацията. Въпреки това в текста на самия стандарт, който е посветен на управлението на данните през целия жизнен цикъл на обектите с помощта на BIM, съкращението BIM се споменава, но никога не е ясно дефинирано

Първоначалният уебсайт на доставчика, който публикува поредица от бели книги за BIM през 2002 г. [60] и 2003 г. [61], всъщност възпроизвежда маркетингови материали за концепциите BOM (Bills of Materials) и PLM (Product Lifecycle Management), използвани преди това в софтуера за машиностроење Pro-E през 90-те години на миналия век [63].

Информационното моделиране на сгради, иновативен нов подход към проектирането, строителството и управлението на сгради, въведен от..... [име на CAD доставчик] през 2002 г., промени начина, по който професионалистите в бранша по целия свят мислят за това как технологията може да се прилага при проектирането, строителството и управлението на сгради.

- BIM Whitepaper, 2003 [61]

Тези ранни публикации свързват BIM директно с концепцията за централизирана интегрирана база данни. Както е посочено в "Whitepaper" от 2003 г., BIM е управление на информацията за сградата,

при което всички актуализации се извършват в единно хранилище, като се гарантира, че всички чертежи, разрези и спецификации (BOM - Bills of Materials) са синхронизирани.

*BIM се описва като управление на строителната информация, при което всички актуализации и промени се извършват в база данни. Така независимо дали става въпрос за схеми, разрези или листови чертежи, всичко е винаги координирано, последователно и актуално.*

- Уебсайт на фирма за CAD продавач с BIM Whitepaper, 2003 г. [54]

Идеята за управление на проектирането чрез единна интегрирана база данни е широко обсъждана още в изследванията през 80-те години на миналия век. Например концепцията за БДС на Чарлз Ийстман [57] включва 43 препратки към термина "база данни" (фигура 6.1-2). До 2004 г. този брой е намалял почти наполовина до 23 в Бялата книга за BIM от 2002 г. [64]. А към средата на 2000 г. темата за базите данни практически изчезна от маркетинговите материали на доставчиците и от програмата за цифровизация като цяло.

Въпреки че първоначално базата данни и достъпът до нея са били замислени като ядро на BIM - системата, с течение на времето акцентът се измества върху геометрията, визуализацията и 3D. Самият регистратор на стандарта IFC през 1994 г., който публикува BIM Whitepaper през 2002 г. - същият доставчик - в Whitepaper от началото на 2000 г. изрично посочва ограниченията на неутралните формати като IGES, STEP и IFC и необходимостта от пряк достъп до CAD бази данни:

Възможно е различните приложения да са несъвместими и повторно въведените данни да са неточни [...]. Резултатът от традиционното проектиране с помощта на компютър [CAD]: по-високи разходи, по-дълъг срок за пускане на пазара и по-ниско качество на продукта. Днес всички основни приложения използват стандартни индустритални интерфейси за обмен на данни на ниско ниво. Като използват старите стандарти IGES или новите STEP [IFC е де факто и де юре копие на формата STEP/IGES] за обмен на данни между приложения от различни производители, потребителите могат да постигнат известна съвместимост на данните между най-добрите продукти. Но IGES и STEP работят само на ниски нива и не могат да обменят данни, които са толкова богати, колкото информацията, генерирана от съвременните водещи приложения [...]. И макар че тези и други стандарти се подобряват почти ежедневно, те винаги ще изоставят от продуктите на днешните производители по отношение на богатството на данните. [...] Програмите в рамките на едно приложение трябва да могат да обменят и запазват богатството на данните, без да прибягват до неутрални транслатори като IGES, STEP [IFC] или PATRAN. Вместо това рамковите приложения трябва да могат да осъществяват директен достъп до основната CAD база данни, така че да не се губят детайлите и точността на информацията.

- Бяла книга на доставчика на CAD (IFC, BIM) "Integrated Design and Manufacturing: Benefits and Rationale" (Интегрирано проектиране и производство: ползи и обосновка), 2000 г. [65]

Така още през 80-те години и в началото на 2000 г. за ключов елемент на цифровото проектиране в средата на CAD се смята базата данни, а не форматът на файла или неутралният формат IFC. Предлагаше се да бъдат изоставени транслаторите и приложенията да имат прям достъп до данните. В действителност обаче към средата на 20-те години на миналия век концепцията за BIM започна да прилича на стратегия "разделяй и владей", при която интересите на производителите на софтуер, използващи затворени геометрични ядра, са приоритетни пред развитието на отворения обмен на информация.

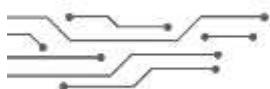
Днес BIM се възприема като неразделна част от строителната индустрия. Но през последните две десетилетия обещанията за опростено сътрудничество и интегриране на данни до голяма степен останаха нереализирани. Повечето решения все още са обвързани със затворени или неутрални формати и специализирани инструменти. Ще разгледаме подробно историята на BIM, отворения BIM и IFC, както и въпросите за оперативната съвместимост и геометричните ядра в част 6 на книгата "CAD и BIM: маркетинг, реалност и бъдещето на проектните данни в строителството".

Днес индустрията е изправена пред ключовото предизвикателство да премине от традиционното разбиране на CAD (BIM) като инструмент за моделиране към използването му като пълноценна база данни. Това изисква нови подходи за работа с информация, изоставяне на зависимостта от затворени екосистеми и прилагане на отворени решения.

С развитието на инструментите за обратно инженерство, които позволяват достъп до базите данни на CAD, както и с разпространението на технологиите с отворен код и LLM, потребителите и разработчиците в строителната индустрия все повече се отдалечават от неясните условия на доставчиците на софтуер. Вместо това фокусът се измества върху това, което наистина има значение: данни (бази данни) и процеси.

Зад модерните съкращения и визуализации се крият стандартни практики за управление на данни: съхранение, прехвърляне и преобразуване - т.е. класическият процес ETL (Extract, Transform, Load). Както и в други отрасли, цифровизацията на строителството изиска не само стандарти за обмен, но и ясно структурирана работа с хетерогенна информация.

За да използват напълно потенциала на CAD (BIM) данните, компаниите трябва да преосмислят подхода си към управлението на информацията. Това неминуемо ще доведе до ключов елемент на цифровата трансформация - унифициране, стандартизиране и смислено структуриране на данните, с които строителните специалисти работят ежедневно.

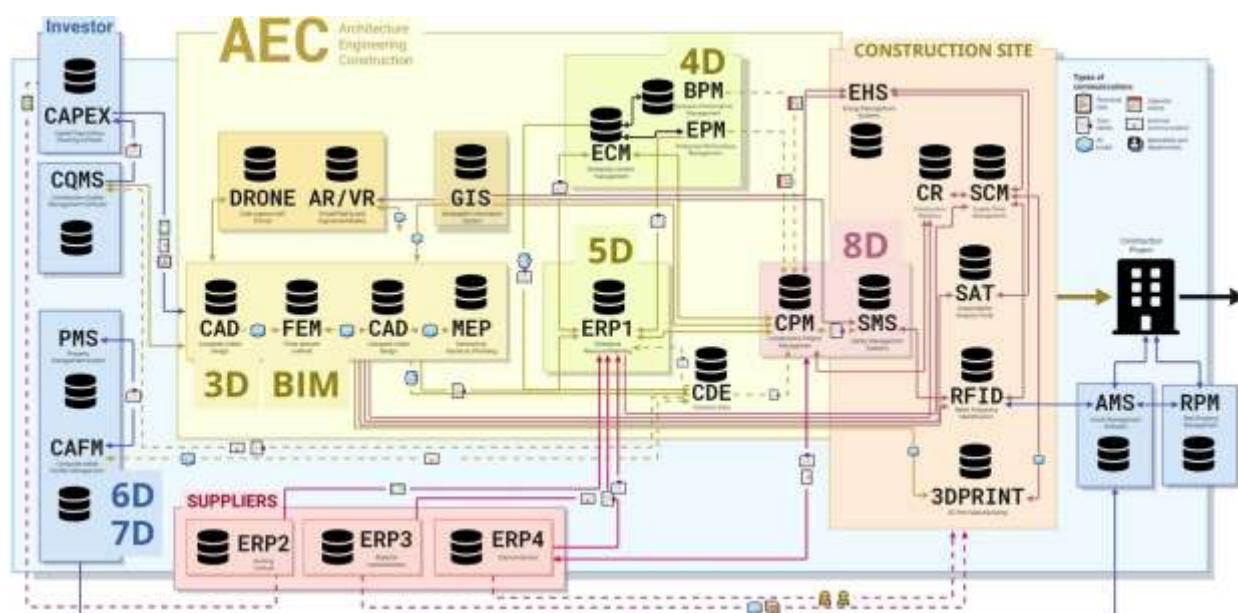


## ГЛАВА 3.2.

### ХАРМОНИЗИРАНЕ И СТРУКТУРИРАНЕ НА ДАННИ

#### Запълване на системите с данни в строителната индустрия

Независимо дали става въпрос за големи корпорации или за средни предприятия, специалистите ежедневно се занимават с попълването на софтуерни системи и бази данни с различни интерфейси с многоформатна информация (фиг. 3.2-1), които с помощта на мениджърите трябва да си сътрудничат. Именно този комплекс от взаимодействащи си системи и процеси в крайна сметка създава приходи и печалба за компанията.



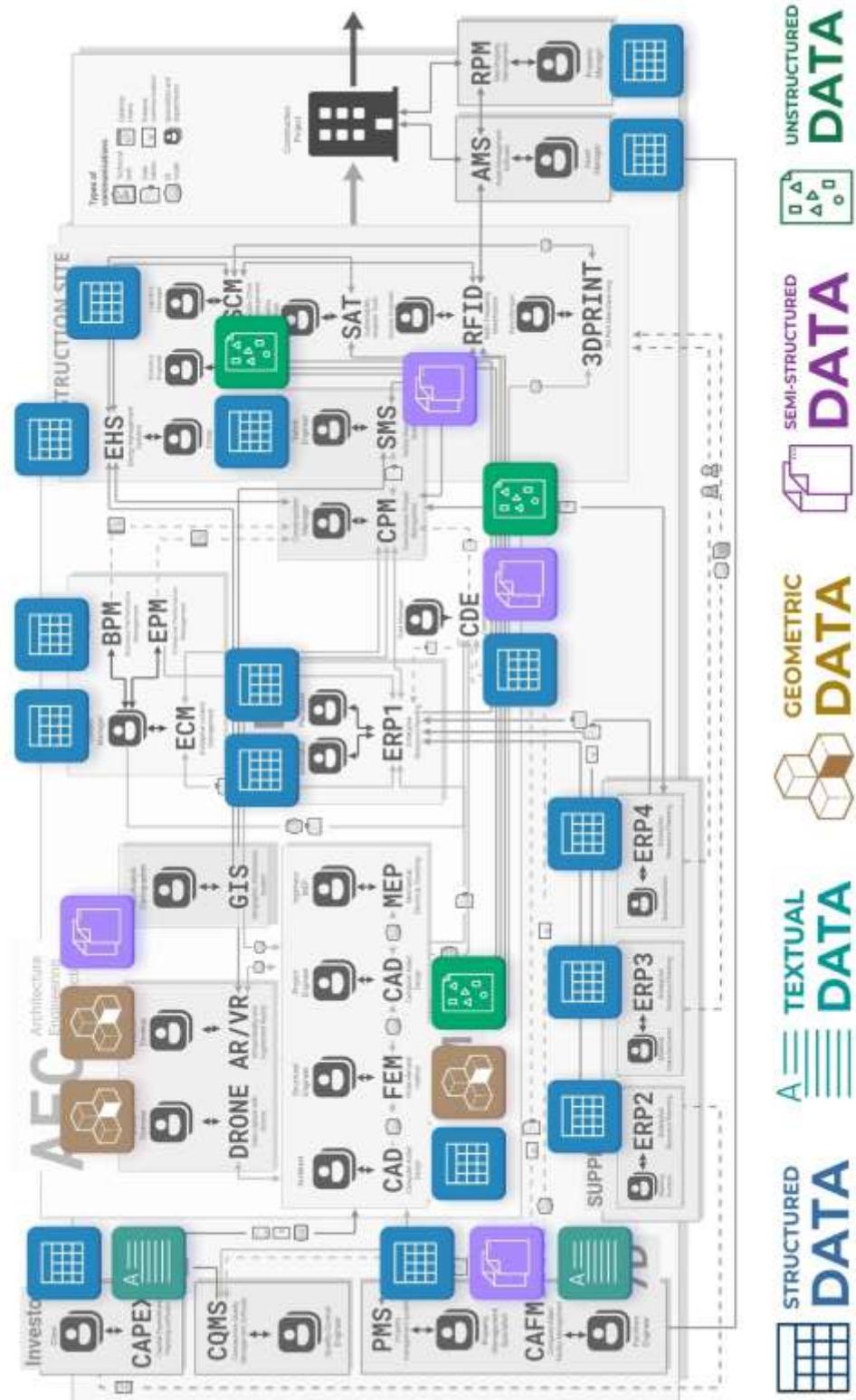
Фигура 3.2-1 На практика всяка система или приложение, използвани в строителния бизнес, имат в основата си една от популярните бази данни RDBMS.

Всяка от категориите системи, споменати по-рано и прилагани в строителния бранш, работи със собствени типове данни, съответстващи на функционалната роля на тези системи. За да преминем от абстрактното ниво към конкретното, преминаваме от типовете данни към тяхното представяне като формати и документи.

Към предоставения по-рано списък на системите (фигура 1.2-4) сега добавяме специфичните видове формати и документи, с които те често работят:

#### ■ Инвеститор (CAPEX)

- Финансови данни: бюджети, прогнози за разходите (структурирани данни).
- Данни за пазарните тенденции: пазарни анализи (структурирани и неструктурирани данни).
- Правни и договорни данни: договори (текстови данни).



Фигура 3.2-2 В строителната индустрия се използват много системи с различни интерфейси, които работят с различни видове данни.

### **Системи за управление (PMS, CAFM, CQMS)**

- Данни за проекта: графики, задачи (структурирани данни).
- Данни за поддръжката на съоръженията: планове за поддръжка (текстови и полуструктурни данни).
- Данни за контрол на качеството: стандарти, доклади от проверки (текстови и неструктурни данни).

### **CAD, FEM и BIM**

- Технически чертежи: архитектурни, структурни планове (геометрични данни, неструктурни данни).
- Изграждане на модели: 3D -модели, данни за материалите (геометрични и полуструктурни данни).
- Инженерни изчисления: анализ на натоварването (структурирани данни).

### **Системи за управление на строителната площадка (EHS, SCM)**

- Данни за безопасност и здраве: протоколи за безопасност (текстови и структурирани данни).
- Данни за веригата на доставки: инвентаризации, поръчки (структурирани данни).
- Ежедневни отчети: работни часове, производителност (структурирани данни).

### **Дрони, AR/VR, ГИС, 3D -принтиране**

- Геоданни: топографски карти (геометрични и структурирани данни).
- Данни в реално време: видео и снимки (неструктурирани данни).
- Модели за 3D -отпечатване: цифрови чертежи (геометрични данни).

### **Допълнителни системи за управление (4D BPM, 5D ERP1)**

- Данни за времето и разходите: графици, оценки (структурирани данни).
- Управление на промените: записи на промените в проекта (текст и структурирани данни).
- Отчитане на изпълнението: показатели за успех (структурирани данни).

### **Интегриране на данни и комуникация (CDE, RFID, AMS, RPM)**

- Обмен на данни: обмен на документи, модели на данни (структурирани и текстови данни).
- RFID и данни за проследяване: логистика, управление на активи (структурирани данни).
- Мониторинг и контрол: сензори на обектите (структурирани и неструктурни данни).

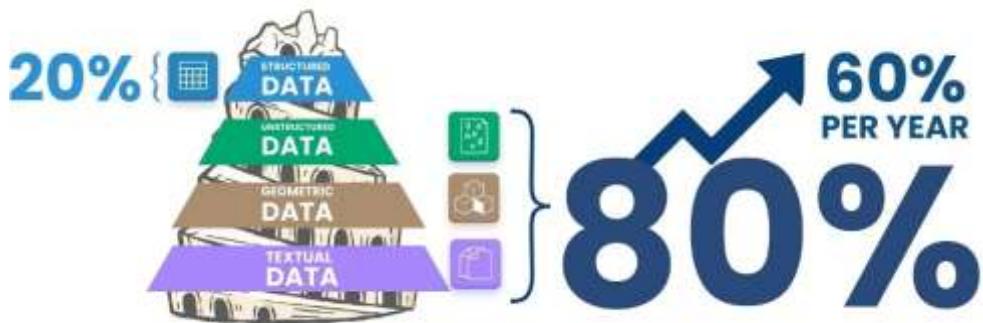
По този начин всяка система в строителната индустрия - от системите за управление на строителни обекти до оперативните бази данни - работи със свой собствен вид информация: структурирана, текстова, геометрична и други. "Пейзажът от данни", с който специалистите трябва да работят ежедневно, е изключително разнообразен. Простото изброяване на форматите обаче не разкрива сложността на реалната работа с информация.

На практика компаниите се сблъскват с факта, че данните, дори когато са извлечени от системите, не са готови да бъдат използвани "както са". Това важи особено за текстове, изображения, PDF файлове, CAD файлове и други формати, които трудно се анализират със стандартни инструменти. Ето защо следващата ключова стъпка е трансформацията на данните - процес, без който обработката, анализът, визуализацията и вземането на решения не могат да бъдат ефективно автоматизирани.

## Трансформация на данните: критичната основа на съвременния бизнес анализ

Днес повечето компании са изправени пред парадокс: около 80% от ежедневните им процеси все още разчитат на класически структурирани данни - познатите електронни таблици на Excel и релационни бази данни (RDBMS) [66]. В същото време обаче 80% от новата информация, която навлиза в цифровата екосистема на компаниите, е неструктуррирана или слабо структурирана (фиг. 3.2-3) [52]. Това включва текст, графики, геометрия, изображения, CAD -модели, документация в PDF, аудио- и видеозаписи, електронна кореспонденция и много други.

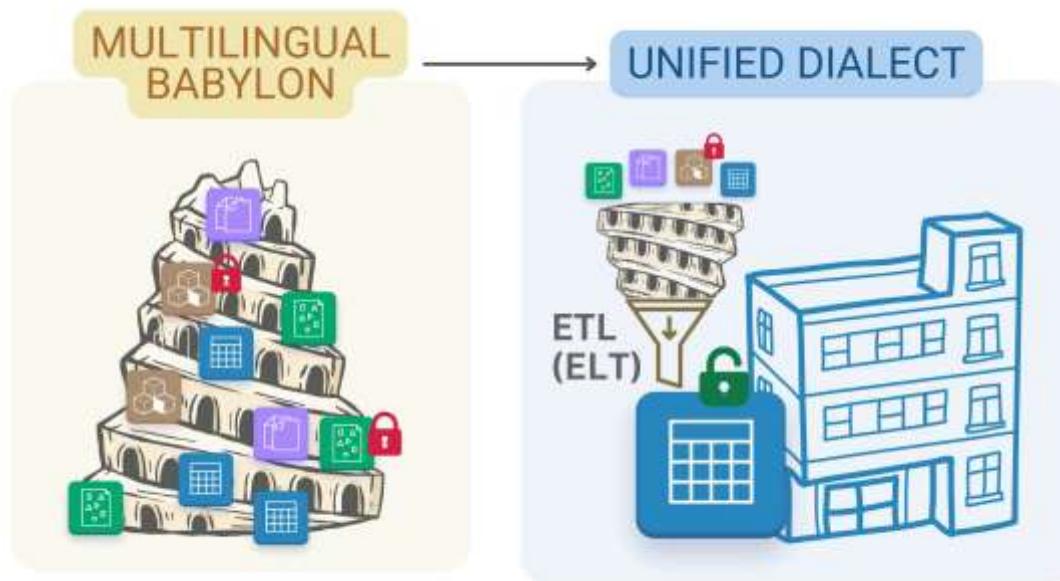
Освен това обемът на неструктуррираните данни продължава да нараства бързо - годишният темп на нарастване се оценява на 55-65% [67]. Подобна динамика създава сериозни трудности при интегрирането на новата информация в съществуващите бизнес процеси. Пренебрегването на този поток от многоформатни данни води до формиране на информационни пропуски и намалява управляемостта на цялата цифрова среда на компанията.



Фиг. 3.2-3 Ежегодното нарастване на неструктуррираните данни създава предизвикателства при интегрирането на поточната информация в бизнес процесите.

Игнорирането на сложни неструктурирани и объркващи слабо структурирани данни в процесите на автоматизация може да доведе до значителни пропуски в информационния пейзаж на компанията. В днешния свят на неконтролирамо и лавинообразно движение на информация компаниите трябва да възприемат хибриден подход към управлението на данни, който включва ефективни методи за работа с всички видове данни.

Ключът към ефективното управление на данни се крие в организирането, структурирането и класифицирането на различни видове данни "Babel" (включително неструктурирани, текстови и геометрични формати, в структурирани или слабо структурирани данни). Този процес превръща хаотичните набори от данни в организирани структури за интегриране в системи, като по този начин позволява вземането на решения въз основа на тях (фигура 3.2-4).



**Фиг. 3.2-4** Основната задача на отделите за управление на данни е да превърнат "Вавилон" от разнообразни и многоформатни данни в структурирана и категоризирана система.

Една от основните пречки пред подобна хармонизация остава ниското ниво на оперативна съвместимост между различните цифрови платформи - "силозите", за които говорихме в предишните глави.

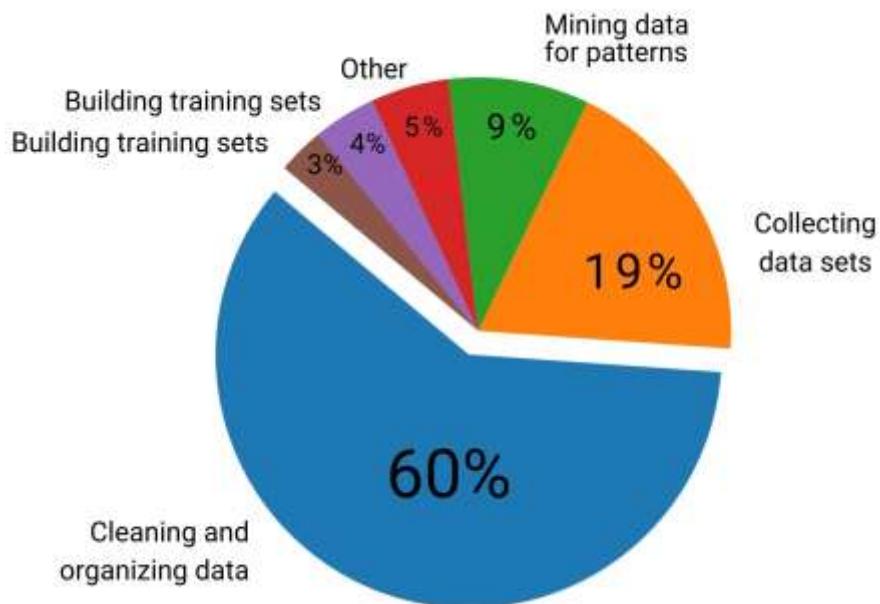
Според доклада Националният институт по стандартизация и технологии (NIST, САЩ) подчертава [68], че лошата съвместимост на данните между различните строителни платформи води до загуба на информация и значителни допълнителни разходи. Само през 2002 г. проблемите с оперативната съвместимост на софтуера са причинили загуби в капиталното строителство в САЩ на обща стойност 15,8 млрд. долара годишно, като две трети от тези загуби се поемат от собствениците и операторите на сгради, особено по време на експлоатацията и поддръжката [68]. В проучването се отбелязва също, че стандартизацията на форматите за данни може да намали тези загуби и да подобри ефективността през целия жизнен цикъл на обекта.

Според проучване на CrowdFlower от 2016 г. [69], обхванало 16 000 учени в областта на данните от цял свят, основният проблем остават "мръсните" и многоформатните данни. Според това проучване най-ценният ресурс не са крайните бази данни или моделите за машинно обучение, а времето, изразходвано за подготовка на информацията.

Почистването, форматирането и организирането отнемат до 60% от времето на анализатора и на мениджъра на данни. Почти една пета се изразходва за търсене и събиране на подходящи набори от данни, които често са скрити в силози и недостъпни за анализ. И само около 9 процента от времето се изразходва директно за моделиране, анализ, изготвяне на прогнози и тестване на хипотези. Останалата част се изразходва за комуникация, визуализация, докладване и проучване на помощни източници на информация.

Средно работата с данни на мениджъра се разпределя по следния начин (фигура 3.2-5):

- **Почистване и организиране на данните (60%):** наличието на чисти и структурирани данни може значително да намали работното време на анализатора на и да ускори процеса на изпълнение на задачите.
- **Събиране на данни (19%):** Основно предизвикателство за специалистите в областта на науката за данните е намирането на подходящи набори от данни. Често фирмени данни са подредени в хаотично организирани "силози", което затруднява достъпа до необходимата информация.
- **Моделиране/машинно обучение (9%):** Често е възпрепятствано от липсата на яснота за бизнес целите от страна на клиентите. Липсата на ясно формулирана мисия може да пропали потенциала дори на най-добрния модел.
- **Други задачи (5%):** освен с обработката на данни анализаторите трябва да се занимават с проучвания, да изследват данните от различни гледни точки, да съобщават резултатите чрез визуализации и доклади и да препоръчват оптимизиране на процеси и стратегии.



**Фиг. 3.2-5 За какво мениджърите на данни, работещи с данни, отделят най-много време (въз основа на [70]).**

Тези оценки се потвърждават и от други проучвания. Според проучване на Xplenty, публикувано в BizReport през 2015 г. [71], между 50% и 90% от времето на специалистите по бизнес интелигентност (BI) (BI) се изразходва за подготовка на данните за анализ.

Почистването, валидирането и организирането на данните от представлява критична основа за всички процеси за обработка на данни и анализ надолу по веригата, като отнема до 90% от времето на специалистите по данни.

Този старателен и невидим за крайния потребител труд е от решаващо значение. Грешките в необработените данни неизбежно изкривяват анализите, подвеждат и могат да доведат до скъпоструващи управленски грешки. Ето защо процесите на почистване и стандартизиране на

данни - от премахване на дублирания и попълване на пропуски до хармонизиране на мерните единици и привеждане в съответствие с общ модел - се превръщат в крайъгълен камък на съвремената цифрова стратегия.

По този начин задълбоченото преобразуване, почистване и стандартизиране на данните не само заемат по-голямата част от времето на специалистите (до 80% от работата с данни), но и определят възможността за ефективното им използване в рамките на съвременните бизнес процеси. Въпреки това, само организацията и почистването на данните не изчерпват задачата за оптимално управление на информационните потоци в предприятието. По време на етапа на организиране и структуриране става въпрос за избор на подходящ модел на данните, който пряко влияе върху удобството и ефективността на работата с информацията в следващите етапи на обработка.

Тъй като данните и бизнес целите са различни, е важно да се разберат характеристиките на моделите на данни и да може да се избере или създаде правилната структура. В зависимост от степента на структуриране и начина, по който се описват връзките между елементите, съществуват три основни модела: структуриран, слабо структуриран и графичен. Всеки от тях е подходящ за различни задачи и има своите силни и слаби страни.

## Модели на данни: връзки в данните и връзки между елементите

Данните в информационните системи се организират по различни начини - в зависимост от задачите и изискванията за съхраняване, обработка и предаване на информация. Основната разлика между видовете модели на данни, формата, в която се съхранява информацията, е степента на структуриране и начинът, по който се описват връзките между елементите.

Структурирани данни имат ясна и повтаряща се схема: те са организирани като таблици с фиксирани колони. Този формат осигурява предсказуемост, лесна обработка и ефективност при извършване на заявки на SQL, филтриране и агрегиране. Примери - бази данни (RDBMS), Excel, CSV.

Свободно структурирани данни позволява гъвкава структура: различните елементи могат да съдържат различен набор от атрибути и да се съхраняват като йерархии. Примери за това са JSON, XML или други формати на документи. Тези данни са удобни, когато е необходимо да се моделират вложени обекти и връзки между тях, но от друга страна, усложняват анализа и стандартизацията на данните (фиг. 3.2-6).

Data Model	Storage Format	Example
	<b>Relational</b>	CSV, SQL A table of doors in Excel
	<b>Hierarchical</b>	JSON, XML Nested door objects inside a room
	<b>Graph-based</b>	RDF, GraphDB Relationships between building elements

Фигура 3.2-6 Моделът на данните е логическа структура, която описва как данните се организират, съхраняват и обработват в дадена система.

Изборът на подходящ формат зависи от целите:

- Ако скоростта на филтриране и анализ е важна - релационни таблици (SQL, CSV, RDBMS, колонарни бази данни) ще са достатъчни.
- Ако е необходима гъвкавост на структурата, по-добре е да използвате JSON или XML.
- Ако данните имат сложни връзки - графичните бази данни осигуряват видимост и мащабируемост.

В класическите релационни бази данни (RDBMS) всяка същност (напр. врата) се представя с ред, а нейните свойства - с колони на таблицата. Например една таблица с елементи от категорията "Врати" може да съдържа полетата ID, Височина, Ширина, Пожароустойчивост и ID на помещението, указващо помещението (фигура 3.2-7).

В класическите релационни бази данни (СУБД) връзките се формират под формата на таблици, в които всеки запис представлява обект, а колоните - неговите параметри. В табличен формат данните за вратите в проекта изглеждат по следния начин, където всеки ред представлява отделен елемент - врата с нейния уникален идентификатор и атрибути, а връзката със стаята се осъществява чрез параметъра "ID на стаята".



Door ID	Room ID	Height (mm)	Width (mm)	Fireproof
ID1001	101	2000	900	Yes
ID1002	101	2100	800	No
ID1003	102	2000	850	Yes

Фиг. 3.2-7 Информация за трите елемента от категорията "Врати" на проекта в табличен структуриран вид.

В свободно структурирани формати, като JSON или XML, данните се съхраняват в йерархична или вложена форма, в която елементите могат да съдържат други обекти, а структурата им може да варира. Това позволява да се моделират сложни връзки между елементите. Подобна информация за вратите в проекта, която е записана в структуриран вид (фигура 3.2-7), е представена в свободно структуриран формат (JSON) по такъв начин (фигура 3.2-8), че те се превръщат във вложени обекти в рамките на стаи (Rooms - ID), което логически отразява йерархията.

```

1 {
2   "Rooms": [
3     {
4       "ID": 101,
5       "Doors": [
6         {"ID": 1, "Height": 2000, "Width": 900, "Fireproof": "Yes"},
7         {"ID": 2, "Height": 2100, "Width": 800, "Fireproof": "No"}
8       ]
9     },
10    {
11      "ID": 102,
12      "Doors": [
13        {"ID": 3, "Height": 2000, "Width": 850, "Fireproof": "Yes"}
14      ]
15    }
16  ]
17 }
18

```

Фиг. 3.2-8 Информация за елементите от категорията "Врати" на проекта в JSON формат.

В модела на граф данните се представят като възли (върхове) и връзки (ребра) между тях. Това ви позволява да визуализирате сложните връзки между обектите и техните атрибути. В случая с данните за врати и стаи в проекта графичното представяне на е следното:

- **Възлите (nodes)** представляват основните обекти: стаи (Room 101, Room 102) и врати (ID1001, ID1002, ID1003).
- **Ребрата (връзките)** показват връзките между тези обекти, например принадлежността на врата към определена стая.
- **Атриутите** са съпоставени с възли и съдържат свойства на обекта (височина, ширина, пожароустойчивост за врати).



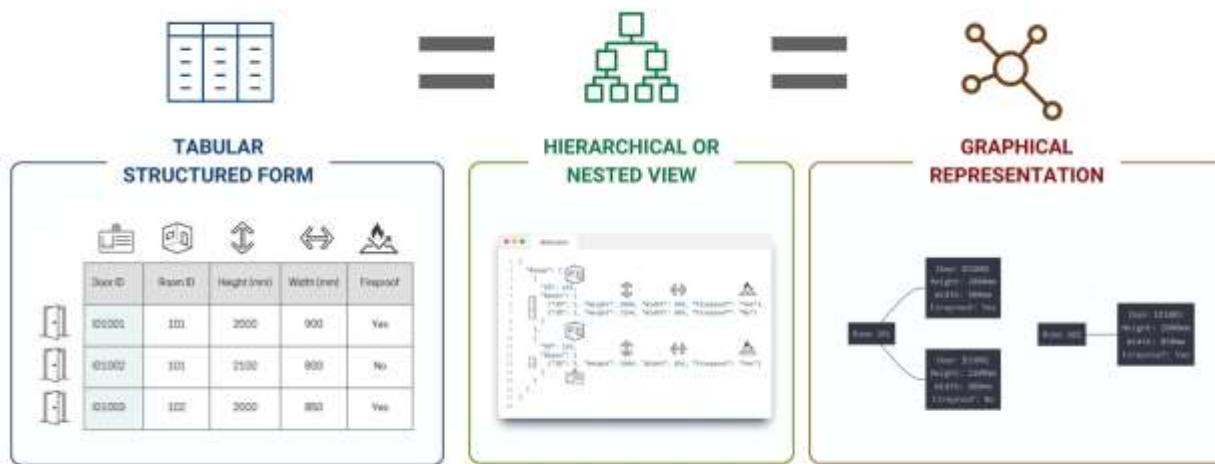
Фиг. 3.2-9 Информация за обекта на вратата на проекта в графичен вид.

В графовия модел на данните за описание на врати всяка стая и всяка врата са отделни възли.

Вратите са свързани със стаите чрез ребра, които показват, че вратата принадлежи към определена стая. Атриутите на вратите (височина, широчина, пожароустойчивост) се съхраняват като свойства на съответните възли. Повече подробности за графовите формати и как графовата семантика се появява в строителството ще бъдат разгледани в главата "Появата на семантика и онтология в строителството".

Базите данни с графики са ефективни, когато са важни връзките между данните, а не самите данни, например в системите за препоръчване, системите за маршрутизация или при моделиране на сложни връзки в проекти за управление на съоръжения. Графовият формат опростява създаването на нови взаимоотношения, като позволява добавянето на нови типове данни към графа, без да се променя структурата на съхранение. Въпреки това в сравнение с релационните таблици и структурираните формати в графа няма допълнителна свързаност на данните - прехвърлянето на двуизмерни данни от база данни в граф не увеличава броя на връзките и не предоставя нова информация.

Формата и схемата на данните трябва да бъдат съобразени с конкретния случай на използване и задачите, които трябва да бъдат изпълнени. За да се работи ефективно в бизнес процесите, е важно да се използват онези инструменти и онези модели на данни, които помагат да се получат резултати възможно най-бързо и лесно.



Фигура 3.2-10 Една и съща информация за елементите на проекта може да се съхранява в различни формати, като се използват различни модели на данни.

Днес повечето големи компании са изправени пред проблема с прекомерната сложност на данните. Всяко от стотиците или хилядите приложения използва свой собствен модел на данните, което създава прекомерна сложност - отделният модел често е десетки пъти по-сложен от необходимото, а съвкупността от всички модели е хиляди пъти по-сложна. Тази прекомерна сложност значително затруднява работата както на разработчиците, така и на крайните потребители.

Тази сложност налага сериозни ограничения върху разработването и поддръжката на системите

на компанията. Всеки нов елемент в модела изисква допълнителен код, внедряване на нова логика, задълбочено тестване и адаптиране към съществуващите решения. Всичко това увеличава разходите и забавя работата на екипа по автоматизация в компанията, като превръща дори прости задачи в скъпи и времеемки процеси.

Сложността засяга всички нива на архитектурата на данните. В релационните бази данни тя се изразява в нарастващия брой таблици и колони, които често са излишни. В обектно-ориентирани системи сложността се увеличава от множеството класове и взаимосвързани свойства. Във формати като XML или JSON сложността се проявява чрез обврквачи вложени структури, уникални ключове и непоследователни схеми.

Прекалената сложност на моделите на данни прави системите не само по-малко ефективни, но и трудни за разбиране от крайните потребители, а в бъдеще и от големите езикови модели и агенти на LLM. Именно проблемът с разбирането и сложността на моделите на данни и обработката на данни повдига въпроса: как да направим данните достатъчно лесни за използване, така че те действително да започнат да бъдат полезни бързо.

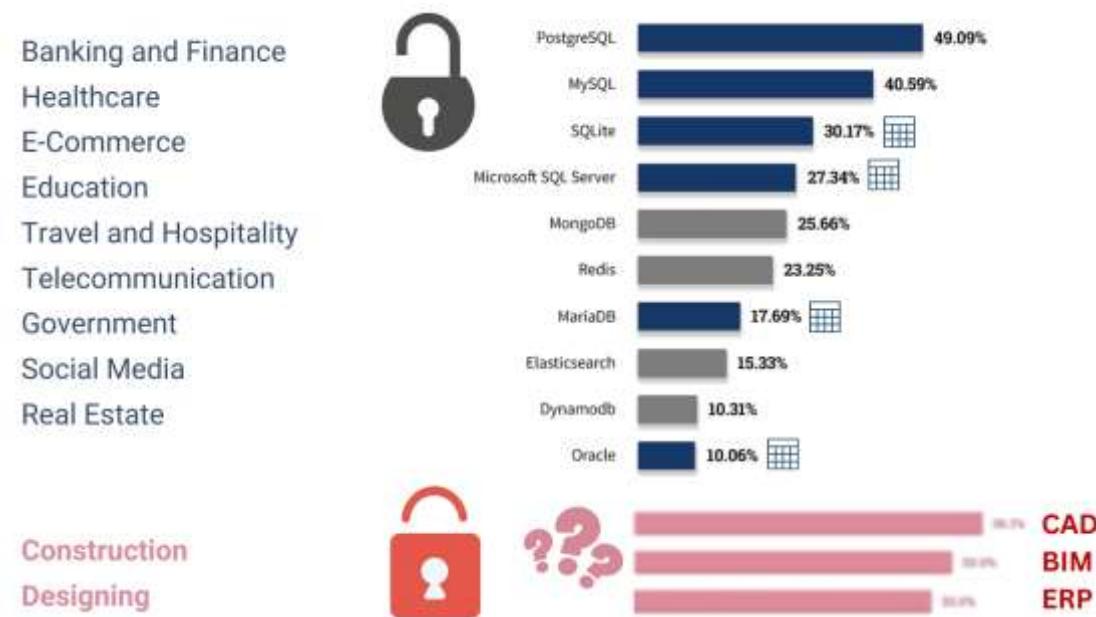
Дори когато моделите на данни са избрани разумно, тяхната полезност намалява значително, ако достъпът до данните е ограничен. Патентованите формати и затворените платформи възпрепятстват интеграцията, усложняват автоматизацията и отнемат контрола върху патентованата информация, като създават не просто силоз от нови данни, а заключен силоз, до който може да се получи достъп само с разрешение на доставчика. За да разберем мащаба на проблема, е важно да разгледаме как точно затворените системи влияят на цифровите процеси в строителството.

## **Запазени формати и тяхното въздействие върху цифровите процеси**

Едно от основните предизвикателства, с които се сблъскват строителните компании по време на цифровизацията, е ограниченият достъп до данни. Това затруднява интегрирането на системите, намалява качеството на информацията и затруднява организирането на ефективни процеси. Използването на патентовани формати и затворени софтуерни решения често е в основата на тези трудности.

За съжаление, досега много програми, използвани в строителната индустрия, позволяват на потребителя да записва данни само в патентовани формати или в облачни хранилища, до които има достъп само чрез строго ограничени интерфейси. И нерядко тези решения са изградени в зависимост от още по-затворени системи на по-големи доставчици. В резултат на това дори тези разработчици, които биха искали да предложат по-отворени архитектури, са принудени да се съобразяват с правилата, диктувани от големите доставчици.

Въпреки че съвременните системи за управление на строителни данни все повече поддържат отворени формати и стандарти (фиг. 3.1-5), базираните на CAD (BIM) бази данни и свързаните с тях ERP и CAFM системи остават изолирани собственически "острови" в цифровия пейзаж на индустрията (фиг. 3.2-11).

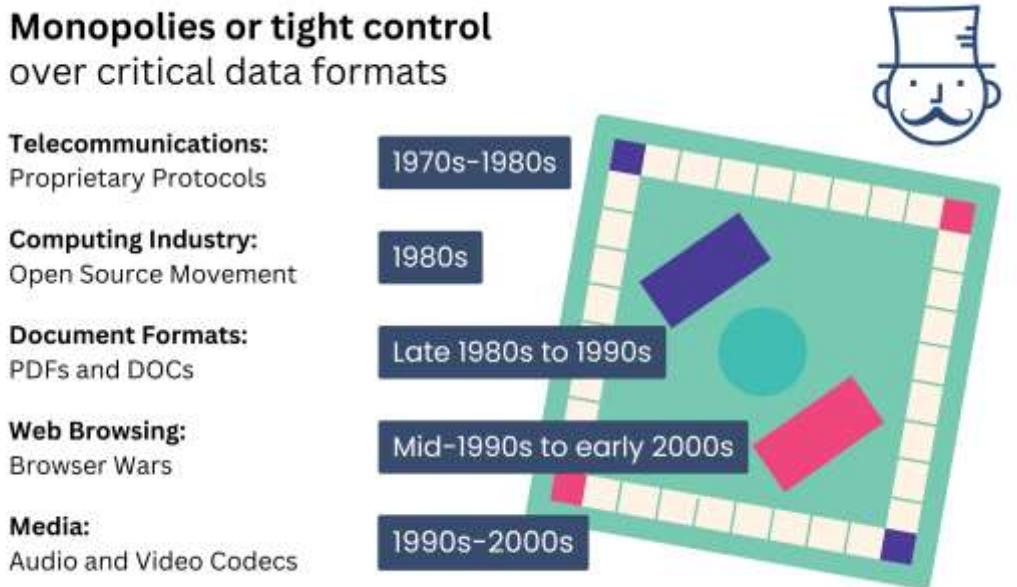


Фигура 3.2-11 Затвореният и патентован характер на данните създава бариери пред интеграцията и достъпа до тях.

Затворените и монополизирани формати и протоколи са проблем не само за строителната индустрия. В много сектори на икономиката борбата със затворените стандарти и ограничения достъп до данни започна със забавяне на иновациите (фигура 3.2-12), наличие на изкуствени бариери за навлизане на нови участници и задълбочаване на зависимостта от големите доставчици. С бързото нарастване на значението на данните органите за защита на конкуренцията просто нямат време да реагират на предизвикателствата, породени от новите цифрови пазари, и в резултат на това затворените формати и ограниченият достъп до данни по същество се превръщат в цифрови "граници", които ограничават потока на информация и растежа [63].

Ако машините произвеждат всичко, от което се нуждаем, тогава положението ни ще зависи от начина, по който тези стоки се разпределят. Всеки ще може да се радва на благодеяние само ако богатството, произведено от машините, се споделя. Или пък повечето хора ще се окажат в крайна сметка в крайна бедност, ако собствениците на автомобили успеят да лобират успешно срещу предразпределението на благата. Засега нещата изглежда се развиват по втория начин, като технологиите водят до все по-голямо неравенство [72].

- Стивън Хокинг, астрофизик, 2015 г.



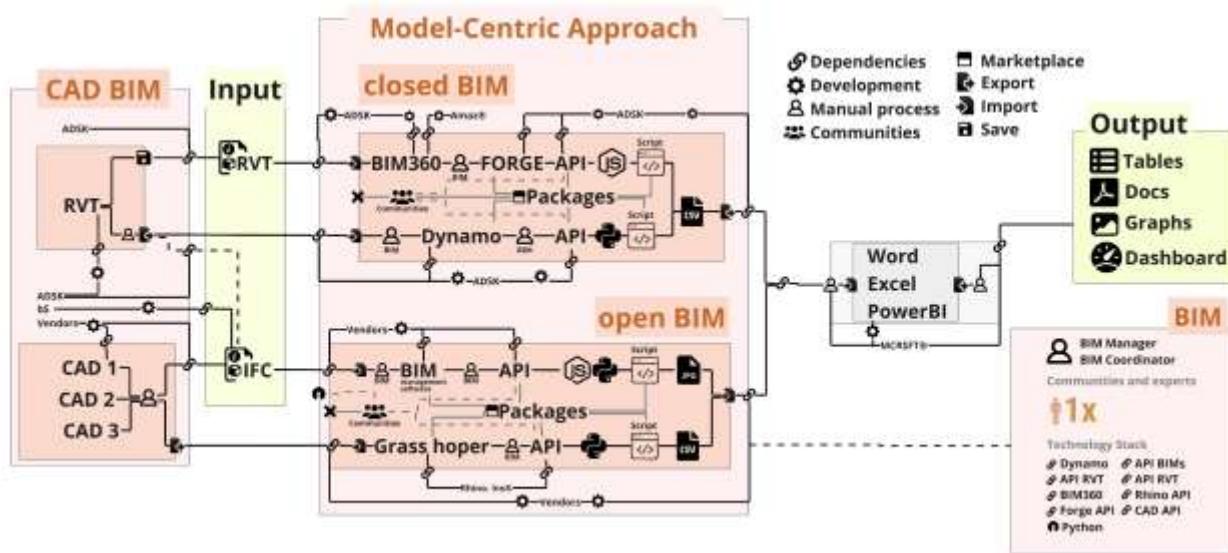
Фиг. 3.2-12 Монополната собственост върху ключови формати и протоколи за данни не е изключителен проблем за строителната индустрия.

В резултат на това, поради затворения достъп до базите данни програмите, мениджърите на данни, анализаторите на данни, ИТ специалистите и разработчиците, създаващи приложения за достъп до данни, обработка и автоматизация в строителната индустрия, днес са изправени пред многобройни зависимости от доставчиците на софтуер (фиг. 3.2-13). Тези зависимости под формата на допълнителни слоеве за достъп изискват създаването на решения със специализирани API-връзки и специални инструменти и софтуер.

API (Application Programming Interface - интерфейс за програмиране на приложения) е формализиран интерфейс, чрез който една програма може да взаимодейства с друга, като обменя данни и функционалност, без да има достъп до изходния код. API описва какви заявки може да прави външната система, в какъв формат трябва да бъдат те и какви отговори ще получи. Това е стандартизиран "договор" между софтуерни модули.

Големият брой зависимости от затворени решения води до това, че цялата архитектура на кода и логиката на бизнес процесите в една компания се превръща в "спагети архитектура" от инструменти, зависещи от политиката на доставчика на софтуер за осигуряване на качествен достъп до данни.

Зависимостта от затворени решения и платформи води не само до загуба на гъвкавост, но и до реални бизнес рискове. Промяна на лицензионните условия, затваряне на достъпа до данни, промяна на форматите или структурата на API - всичко това може да блокира критични процеси. Изведенъж се оказва, че актуализирането на една таблица изисква преработване на цял блок от интеграции и конектори (фиг. 3.2-13) и всяко мащабно обновяване на софтуера или на неговия доставчик на API се превръща в потенциална заплаха за стабилността на цялата система на компанията.



Фиг. 3.2-13 Пример за големия брой зависимости при обработката на CAD -данни създава пречки пред интеграцията на данни в екосистемата на строителните компании.

Разработчиците и системните архитекти в такива условия са принудени да работят не за очакване, а за оцеляване. Вместо да прилагат нови решения, те се адаптират. Вместо да разработват, те се опитват да поддържат съвместимост. Вместо да автоматизират и ускоряват процесите, те прекарват времето си в изучаване на поредните затворени интерфейси, документация за API и безкрайно преработване на кода.

Работата със затворени формати и системи не е просто техническо предизвикателство, а стратегическо ограничение. Въпреки очевидните възможности, предлагани от съвременната автоматизация, изкуствения интелект, LLM и прогнозния анализ, много компании не успяват да реализират пълния им потенциал. А барierите, издигнати от патентованите формати (фигура 3.2-13), лишават предприятията от достъп до собствените им данни. Това може да е иронията на цифровата трансформация в строителството.

Прозрачността на данните и отворените системи не са лукс, а предпоставка за бързина и ефективност. Без откритост бизнес процесите са изпълнени с ненужна бюрокрация, многопластови вериги на одобрение и нарастваща зависимост от принципа HiPPO - вземане на решения въз основа на мнението на най-високоплатения човек.

Въпреки това на хоризонта се очертава промяна на парадигмата. Въпреки господството на патентованите решения, все повече компании осъзнават ограниченията на архитектурите, вдъхновени от Четвъртата индустриска революция. Днес векторът се извества към принципите на Петата революция, където данните като стратегически актив, отворените интерфейси (API) и истинската оперативна съвместимост между системите са в центъра.

Този преход бележи преминаване от затворени екосистеми към гъвкави, модулни цифрови архитектури, в които отворените формати, стандартите и прозрачният обмен на данни са от ключово

значение.

## Отворените формати променят подхода към цифровизацията

Строителният бранш беше един от последните, които се заеха с проблема на затворените и патентовани данни. За разлика от други сектори на икономиката, тук цифровизацията се развива бавно. Причините за това са традиционният консервативен характер на отрасъла, разпространението на разнородни локални решения и дълбоко вкорененият характер на управлението на документи на хартиен носител. В продължение на десетилетия ключовите процеси в строителството разчитаха на физически чертежи, телефонни разговори и несинхронизирани бази данни. В този контекст затворените формати отдавна се възприемат като норма, а не като пречка.

Опитът от други отрасли показва, че премахването на пречките пред затворените данни води до рязко нарастване на иновациите, ускоряване на развитието и засилване на конкуренцията [73]. В науката обменът на отворени данни позволява да се ускорят откритията и да се развие международното сътрудничество. В медицината той може да подобри ефективността на диагностиката и лечението. В софтуерното инженерство - за създаване на екосистеми за съвместно създаване и бързо усъвършенстване на продуктите.

Според доклада на McKinsey "Отворени данни: Отключване на иновациите и производителността с помощта на информационния поток" 2013 г. [74], отворените данни имат потенциала да отключат от 3 до 5 трилиона долара годишно в седем ключови индустрии, включително строителство, транспорт, здравеопазване и енергетика. Според същото проучване децентрализираните екосистеми от данни позволяват на големите строителни компании и изпълнители да намалят разходите за разработване и поддръжка на софтуер, ускорявайки въвеждането на цифрови технологии.

Преходът към отворени архитектури, който отдавна е започнал в други сектори на икономиката, постепенно се разпространява и в строителната индустрия. Големите компании и публичните клиенти, и особено финансовите организации, които контролират инвестициите в строителни проекти, все по-често изискват използването на отворени данни и достъп до изходния код на изчисленията, калкулациите и приложенията. От разработчиците вече не се очаква само да създават цифрови решения и да показват крайните цифри на даден проект - от тях се очаква да бъдат прозрачни, възпроизвеждани и независими от доставчици на приложения от трети страни.

Използването на решения с отворен код дава на клиента увереност, че дори ако външни разработчици спрат да си сътрудничат или напуснат проекта, това няма да се отрази на възможността за по-нататъшно разработване на инструменти и системи. Едно от основните предимства на отворените данни е способността им да елиминират зависимостта на разработчиците на приложения от конкретни платформи за достъп до данни.

Ако компанията не може напълно да се откаже от патентованите решения, възможен компромис е използването на техники за обратно инженерство. Тези законосъобразни и технически издържани методи позволяват затворените формати да бъдат преобразувани в по-достъпни, структурирани и подходящи за интегриране. Това е особено важно, когато има нужда от свързване с

наследени системи или миграране на информация от един софтуерен пейзаж към друг.

Един от най-ярките примери в историята на прехода към отворени формати и прилагането на обратното инженерство (законно хакване на патентовани системи) в строителството е историята на борбата за отваряне на формата DWG, широко използвана в системите за компютърно проектиране (CAD). През 1998 г., в отговор на монопола на един доставчик на софтуер, останалите 15 доставчици на CAD формират нов алианс, наречен "Open DWG", за да предоставят на разработчиците свободни и независими инструменти за работа с формата DWG (фактически стандарт за прехвърляне на чертежи), без да е необходим патентован софтуер или затворени API. Това събитие беше повратна точка, което позволи на десетки хиляди компании да получат свободен достъп до затворения формат на популярно CAD решение от края на 80-те години на миналия век до днес и да създадат съвместими решения, което настъпи конкуренцията на CAD пазара [75]. Днес SDK "Open DWG", който е създаден за първи път още през 1996 г., се използва в почти всички решения, в които е възможно да се импортира, редактира и експортира DWG формат, извън официалното приложение на разработчика на DWG формата.

Други технологични гиганти налагат подобни трансформации. Microsoft, някога символ на патентован подход, отвори изходния код на .NET Framework, започна да използва Linux в инфраструктурата на облачните услуги Azure и придоби GitHub, за да засили позициите си в общността на отворения код. [76]. Meta (бивша Facebook) пусна модели на ИИ с отворен код, като например серията Llama, за да настъпи иновациите и сътрудничеството в разработването на агенти на ИИ. Главният изпълнителен директор Марк Зукърбърг предвижда, че платформите с отворен код ще бъдат водещи в технологичния напредък през следващото десетилетие [77].

Отворен код е модел за разработване и разпространение на софтуер, при който изходният код е отворен за свободно използване, изучаване, модифициране и разпространение.

Отворените данни и решенията с отворен код се превръщат не просто в тенденция, а в основа на цифровата устойчивост. Те дават на компаниите гъвкавост, устойчивост, контрол върху собствените им решения и възможност за мащабиране на цифровите процеси, без да зависят от политиките на доставчиците. И, което е също толкова важно, те връщат на предприятията контрола върху най-ценния ресурс на 21-ви век - техните данни.

## **Смяна на парадигмата: отвореният код като край на ерата на господство на доставчиците на софтуер**

Строителният бранш преживява промяна, която не може да бъде оценена по обичайния начин. Концепцията за подход, ориентиран към данните, и използването на инструменти с отворен код води до преосмисляне на правилата на играта, върху които стъпват софтуерните гиганти на пазара.

За разлика от предишни технологични трансформации, този преход няма да бъде активно настърчаван от доставчиците. Промяната на парадигмата застрашава техните традиционни бизнес модели, основани на лицензиране, абонаменти и консултации. Новата реалност не включва готов продукт или платен абонамент - тя изисква пренастройване на процесите и мисленето.

За да управляват и разработват решения за центрове за данни, базирани на отворени технологии, компаниите ще трябва да преосмислят вътрешните си процеси. Специалистите от различни отдели ще трябва не само да си сътрудничат, но и да преосмислят начина, по който работят заедно.

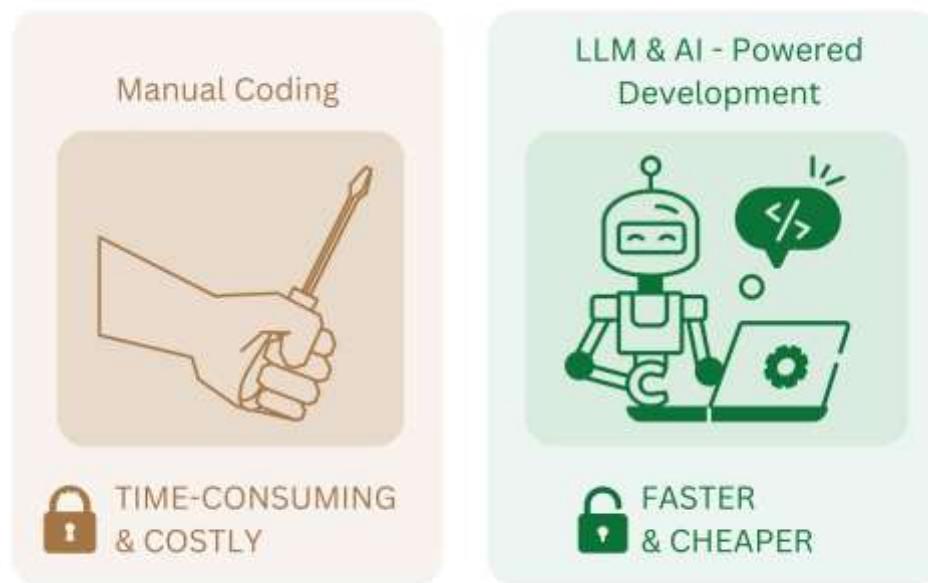
Новата парадигма предполага използването на отворени данни и решения с отворен код, където инструменти, базирани на изкуствен интелект и големи езикови модели (LLM), а не програмисти, ще играят особена роля в създаването на софтуерен код. Още към средата на 2024 г. повече от 25% от новия код в Google ще бъде създаден с помощта на изкуствен интелект [78]. В бъдеще кодирането с LLM ще извърши 80% от работата само за 20% от времето (фигура 3.2-14).

Според проучването на McKinsey за 2020 г. [79] графичните процесори все повече заместват централните процесори в анализа поради високата им производителност и поддръжката им от съвременни инструменти с отворен код. Това позволява на компаниите да ускорят обработката на данни без значителни инвестиции в скъп софтуер или наемане на дефицитни специалисти.

Водещи консултантски фирми, като McKinsey, PwC и Deloitte, подчертават нарастващото значение на отворените стандарти и приложенията с отворен код в различните отрасли.

Според доклада на PwC Open Source Monitor 2019 [80] 69% от компаниите със 100 или повече служители съзнателно използват решения с отворен код. OSS се използва особено активно в големите компании: 71% от компаниите с 200-499 служители, 78% в категорията 500-1999 служители и до 86% сред компаниите с над 2000 служители. Според доклада Synopsys OSSRA 2023 96% от анализираните кодови бази съдържат компоненти с отворен код [81].

Бъдещата роля на разработчика не е да пише ръчно код, а да проектира модели на данни, архитектури на потоци и да управлява агенти с изкуствен интелект, които създават правилните изчисления при поискване. Потребителските интерфейси ще станат минималистични, а взаимодействието ще се основава на диалог. Класическото програмиране ще отстъпи място на проектирането на високо ниво и оркестрирането на цифрови решения (фигура 3.2-14). Съвременните тенденции - като например платформите с нисък код (фиг. 7.4-6) и екосистемите с поддръжка на LLM (фиг. 7.4-4) - ще намалят значително разходите за разработване и поддръжка на ИТ системи.



**Фиг. 3.2-14** Докато днес приложенията се създават ръчно от програмисти, в бъдеще значителна част от кода ще се генерира от решения, базирани на изкуствен интелект и LLM.

Този преход ще бъде различен от предишните и е малко вероятно големите доставчици на софтуер да бъдат катализаторите.

В проучването на Harvard Business School "The Value of Open Source Software" 2024 [40] общата стойност на софтуера с отворен код се оценява от две гледни точки. От една страна, ако изчислим колко би било необходимо, за да се изградят всички съществуващи решения с отворен код от нулата, сумата ще бъде около 4,15 млрд. долара. От друга страна, ако си представим, че всяка компания сама разработва свои собствени аналоги на решенията с отворен код (което се случва навсякъде), без да има достъп до съществуващите инструменти, то общата стойност на бизнеса би достигнала колосалните 8,8 трилиона долара - това е цената на търсенето.

Не е трудно да се предположи, че никой от големите доставчици на софтуер не е заинтересован от свиването на пазар на софтуер с потенциална стойност от 8,8 трилиона долара до едва 4,15 милиарда долара. Това би означавало да се намали обемът на търсенето повече от 2000 пъти. Подобна трансформация е просто нерентабилна за доставчиците, чиито бизнес модели са изградени върху години на поддържане на зависимостта на клиентите от затворени решения. Така че компаниите, които очакват някой да им предложи удобно и отворено решение "до ключ", може да бъдат разочаровани - тези доставчици просто няма да се появят.

Преминаването към отворена цифрова архитектура не означава загуба на работни места или приходи. Напротив, тя създава условия за гъвкави и адаптивни бизнес модели, които в крайна сметка могат да изместят традиционния пазар на лицензи и софтуерни кутии.

Вместо продажба на лицензи - услуги, вместо затворени формати - отворени платформи, вместо

зависимост от доставчик - независимост и възможност за изграждане на решения за реални нужди. Тези, които досега просто са използвали инструменти, ще могат да станат техни съавтори. А тези, които могат да работят с данни, модели, сценарии и логика, ще се окажат в центъра на новата цифрова икономика на индустрията. Ще поговорим повече за тези промени и за това какви нови роли, бизнес модели и формати на сътрудничество се появяват около отворените данни в последната, десета част на книгата.

Решенията, базирани на отворени данни и отворен код, ще позволят на компаниите да се съсредоточат върху ефективността на бизнес процесите, вместо да се борят с останали API-та и да интегрират затворени системи. Съзнателният преход към отворена архитектура може значително да подобри производителността и да намали зависимостта от доставчиците.

Преходът към новата реалност не е само промяна в подходите към разработването на софтуер, но и преосмисляне на самия принцип на работа с данни. В центъра на тази трансформация е не кодът, а информацията: нейната структура, достъпност и интерпретируемост. И именно тук на преден план излизат отворените и структурирани данни, които се превръщат в неразделна част от новата цифрова архитектура.

## Структурирани отворени данни: основата на цифровата трансформация

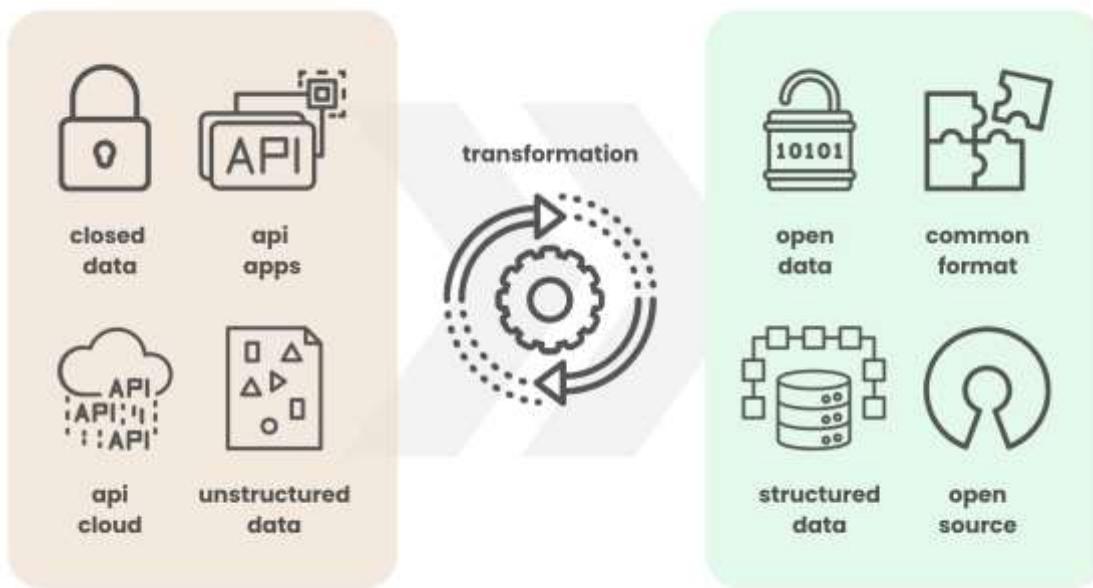
Докато през последните десетилетия устойчивостта на бизнеса се определяше до голяма степен от избора на софтуерни решения и зависимостта от конкретни доставчици, в днешната цифрова икономика ключов фактор е качеството на данните и способността за ефективна работа с тях. Кодът с отворен код е важна част от новата технологична парадигма, но потенциалът му се отключва истински само когато данните са разбираеми, организирани и машинно четими. Сред всички видове модели на данни, структурираните отворени данни се превръщат в крайъгълен камък на устойчивата цифрова трансформация.

Основното предимство на структурираните отворени данни е недвусмисленото им тълкуване и възможността за автоматизирана обработка. Това позволява значително повишаване на ефективността както на ниво отделни операции, така и в цялата организация.

Според доклада на Делойт "Процесът на прехвърляне на данни при трансформацията на предприятието" [82], работата с ИТ за управление на трансфера на структурирани данни е от решаващо значение. Според доклада на правителството на Обединеното кралство "Анализ на данни и изкуствен интелект при изпълнението на правителствени проекти" (2024 г.) [83] премахването на пречките пред обмена на данни между различни проекти и организации е от ключово значение за подобряване на ефективността при управлението на проекти. В документа се подчертава, че стандартизацията на форматите на данните и въвеждането на принципите на отворените данни могат да избегнат дублирането на информация, да сведат до минимум загубата на време и да подобрят точността на прогнозите.

За строителния бранш, в който традиционно преобладава висока степен на фрагментация и разнобразие на форматите, структурираният процес на обединяване и структурираните отворени данни

играят решаваща роля за оформянето на съгласувани и управляеми процеси (фиг. 4.1-14). Те позволяват на участниците в проекта да се съсредоточат върху подобряване на производителността, а не върху решаването на технически проблеми, свързани с несъвместимостта между затворените платформи, моделите и форматите на данни.



**Фиг. 3.2-15 Отворените структурирани данни намаляват зависимостта от софтуерни решения и платформи и ускоряват иновациите.**

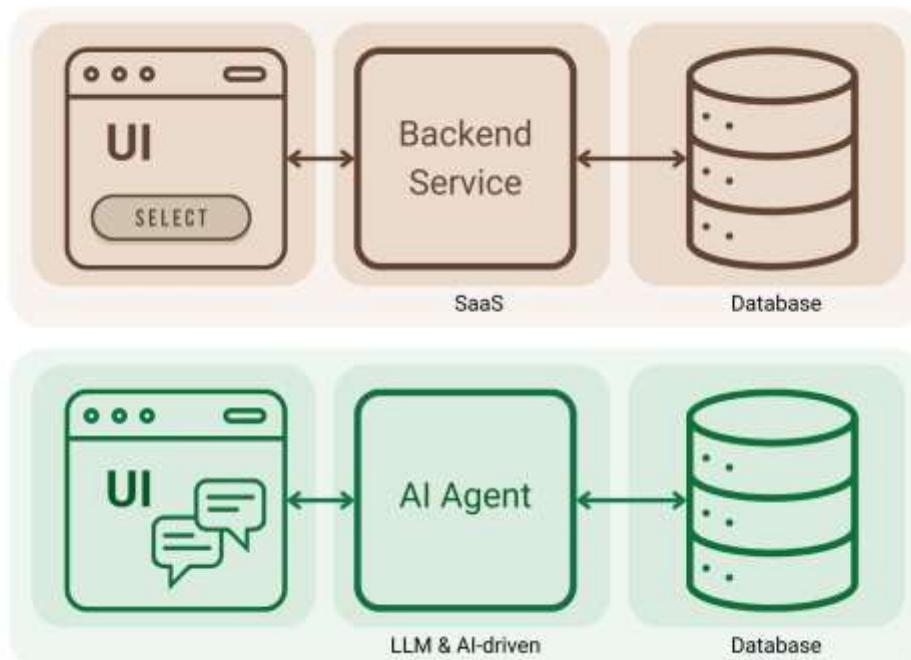
Съвременните технологични инструменти, които ще разгледаме подробно по-нататък в книгата, позволяват не само да се събира информация, но и да се почиства автоматично: да се елиминират дублиранията, да се коригират грешките и да се нормализират стойностите. Това означава, че анализаторите и инженерите не работят с разнородни документи, а с организирана база от знания, подходяща за анализ, автоматизация и вземане на решения.

*Направете го възможно най-просто, но не по-просто.*

- Алберт Айнщайн, физик-теоретик (авторството на цитата се оспорва [84])

Днес повечето потребителски интерфейси за работа с данни могат да се създават автоматично - без да е необходимо ръчно да се пише код за всеки бизнес случай. Това изисква инфраструктурен слой, който разбира структурата, модела и логиката на данните без допълнителни инструкции (фигура 4.1-15). Именно структурата на данните прави този подход възможен: формуляри, таблици, филтри и изгледи могат да се генерират автоматично с минимални усилия за програмиране.

Най-важните интерфейси, важни за потребителя, все още могат да изискват ръчно усъвършенстване. Но в повечето случаи - а това са между 50 и 90% от работните сценарии - е достатъчно автоматичното генериране на приложения и изчисления без използването на специални приложения за тази цел (фиг. 3.2-16), което значително намалява разходите за разработка и поддръжка, намалява грешките и ускорява внедряването на цифрови решения.



Фиг. 3.2-16 Архитектурни модели за работа с данни: традиционна архитектура на приложение и AI -ориентиран модел с LLM.

Преминаването от архитектури, изградени на базата на отделни приложения, към интелигентно управлявани системи, базирани на езикови модели (LLM), е следващата стъпка в цифровата еволюция. В такава архитектура структурираните данни се превръщат не само в обект за съхранение, но и в основа за взаимодействие с инструменти на изкуствения интелект, способни да анализират, тълкуват и препоръчват действия въз основа на контекста.

В следващите глави ще разгледаме реални примери за прилагане на архитектура, базирана на отворени структурирани данни, и ще покажем как езиковите модели се прилагат за автоматично интерпретиране, валидиране и обработка на данни. Тези практически случаи ще ви помогнат да разберете по-добре как работи новата цифрова логика в действие - и какви ползи носи тя на компаниите, които са готови за трансформация.



## ГЛАВА 3.3.

### LLM И ТЯХНАТА РОЛЯ В ОБРАБОТКАТА НА ДАННИ И БИЗНЕС ПРОЦЕСИТЕ

#### **Чат стаи на LLM: ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok за автоматизиране на процесите на данни.**

Появата на големи езикови модели (LLM) е естествено продължение на движението към структурирани отворени данни и философията на отворения код. Когато данните станат организирани, достъпни и машинно четими, следващата стъпка е инструмент, който може да взаимодейства с тази информация, без да е необходимо да се пише сложен код или да се притежават специализирани технически познания.

LLM са пряк продукт на отвореността: големи отворени масиви от данни, публикации и движението за отворен код. Без отворени научни статии, публично достъпни текстови данни и култура на съвместна разработка нямаше да има ChatGPT или други LLM. В известен смисъл LLM е "дестилат" на натрупаното цифрово знание на човечеството, събрано и образовано чрез принципите на откритостта.

Съвременните големи езикови модели (LLM - Large Language Models) като ChatGPT ® (OpenAI), LLaMa™ (Meta AI), Mistral DeepSeek™, Grok™ (xAI), Claude™ (Anthropic), QWEN™ предоставят на потребителите възможността да формулират заявки към данни на естествен език. Това прави работата с информация достъпна не само за разработчици, но и за анализатори, инженери, проектанти, менеджъри и други специалисти, които досега са били далеч от програмирането

**LLM** (Large Language Model) е изкуствен интелект, който е обучен да разбира и генерира текст въз основа на огромни количества данни, събрани от целия интернет. Той е в състояние да анализира контекста, да отговаря на въпроси, да участва в диалог, да пише текстове и да генерира софтуерен код.

Ако по-рано визуализацията, обработката или анализът на данни изискваха познаване на специален език за програмиране: Python, SQL, R или Scala, както и способност за работа с библиотеки като Pandas, Polars или DuckDB и много други, то от 2023 г. ситуацията се променя коренно. Сега потребителят може просто да опише какво иска да получи - и моделът сам ще генерира кода, ще го изпълни, ще покаже таблица или графика и ще обясни резултата. За първи път от десетилетия насам развитието на технологиите не е поело по пътя на усложняването, а на радикалното опростяване и достъпност.

Този принцип - "обработване на данни с думи (подкани)". - бележи нов етап в еволюцията на работата с информация, като на практика извежда създаването на решения на още по-високо ниво на абстракция. Точно както някога вече не беше необходимо потребителите да разбират техническите основи на интернет, за да управляват онлайн магазини или да създават уебсайтове с помощта на WordPress, Joomla и други модулни системи с отворен код (автор книгата работи с такива системи

от 2005 г., включително с образователни и инженерни онлайн платформи). - Това от своя страна доведе до бум на цифровото съдържание и онлайн бизнеса - днес инженери, анализатори и мениджъри могат да автоматизират работните процеси, без да познават езици за програмиране. Това се улеснява от мощните LLM - както безплатни, така и с отворен код, като LLaMA, Mistral, Qwen, DeepSeek и други - които правят напредналите технологии достъпни за възможно най-широва аудитория.

## Големи езикови модели LLM: как работи

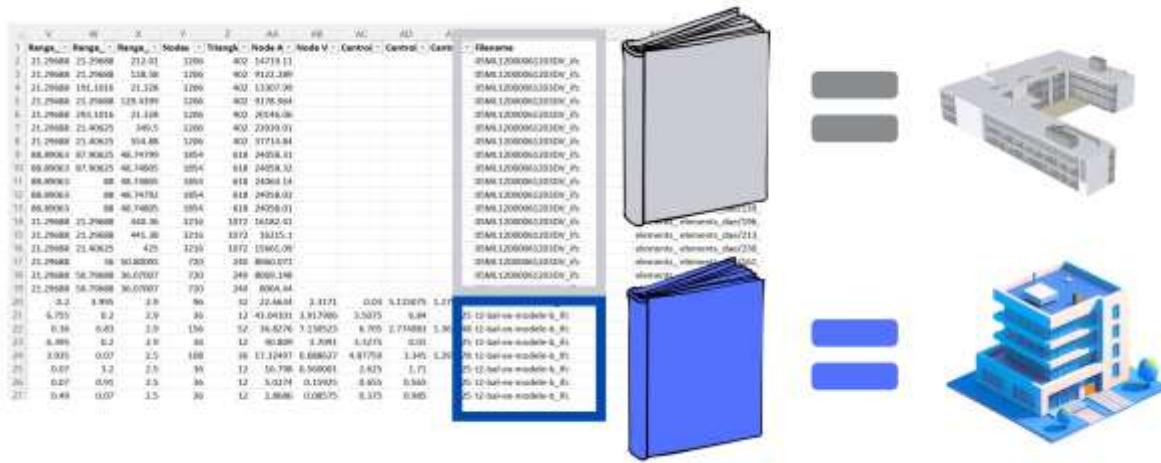
Големите езикови модели (ChatGPT, LLaMA, Mistral, Claude, DeepSeek, QWEN, Grok) са невронни мрежи, обучени върху огромни количества текстови данни от интернет, книги, статии и други източници. Тяхната основна задача е да разбират контекста на човешката реч и да генерират смислени отговори.

Съвременният LLM се основава на архитектурата Transformer, предложена от изследователи на Google през 2017 г. [85]. Ключовият компонент на тази архитектура е механизъмът за внимание, който позволява на модела да отчита връзките между думите, независимо от тяхната позиция в текста.

Процесът на учене на LLM е много подобен на начина, по който хората учат език, само че милиони пъти по-голям. Моделът анализира милиарди примери на думи и изрази, като идентифицира модели в структурата на езика и в логиката на семантичните преходи. Целият текст се разбива на токени - минимални семантични единици (думи или техни части), които след това се трансформират във вектори в многомерно пространство (фиг. 8.2-2). Тези векторни представления позволяват на машината да "разбере" скритите връзки между понятията, а не просто да оперира с текста като с последователност от символи.

Големите езикови модели не са просто инструменти за генериране на текст. Те са в състояние да разпознават смисъла, да намират връзки между понятията и да работят с данни, дори ако те са представени в различни формати. Основното е, че информацията трябва да бъде разбита на разбирами модели и представена като токени, с които LLM да може да работи.

Същият подход може да се приложи и към строителните проекти. Ако мислим за проекта като за вид текст, в който всяка сграда, елемент или конструкция е символ, можем да започнем да обработваме тази информация по подобен начин. Строителните проекти могат да бъдат сравнени с книги, които са организирани в категории, глави и групи от параграфи, състоящи се от минимални токени - елементи на строителния проект (фигура 3.3-1). Като превеждаме моделите на данни в структуриран формат, можем също така да преведем структурираните данни във векторни бази (фигура 8.2-2), които са идеален източник за машинно обучение и технологии като LLM.



**Фиг. 3.3-1 Елементът на строителния проект е като символ в текст: минимална единица, от която се формират групи (параграфи) от раздели (категории) на целия проект.**

Ако строителният проект е дигитализиран и елементите му са представени като токени или вектори, достъпът до тях става възможен на естествен език, а не чрез строги формални заявки. Тук се проявява едно от основните предимства на LLM - способността да се разбере значението на заявката и да се свърже със съответните данни.

Инженерът вече не трябва да пише SQL -запитване или Python код, за да получи необходимите данни - той може просто, разбирайки LLM и структурата на данните, да формулира задачата по обичайния начин: "Намерете всички стоманобетонни конструкции с клас на бетона по-висок от B30 и изчислете общия им обем". Моделът ще разпознае смисъла на заявката, ще я превърне в машинночитаема форма, ще намери данните (ще ги групира и трансформира) и ще върне крайния резултат.

Документите, таблиците, моделите на проектите се преобразуват във векторни представления (вграждане) и се съхраняват в базата данни. Когато потребителят зададе въпрос, заявката също се преобразува във вектор и системата намира най-подходящите данни. Това позволява на LLM да разчита не само на собствените си обучени знания, но и на реални корпоративни данни, дори ако те вече са се появили след края на обучението на модела.

Едно от най-важните предимства на LLM в строителството е възможността за генериране на програмен код. Вместо да предават техническата задача на програмист, специалистите могат да опишат задачата на естествен език, а моделът ще създаде необходимия код, който може да се използва (като се копира от чата) при създаването на код за автоматизация на процеса. LLM-моделите позволяват на специалисти без задълбочени познания по програмиране да допринесат за автоматизацията и подобряването на бизнес процесите в компанията.



**Фигура 3.3-2 LLMs предоставя на потребителите възможността да пишат код и да получават резултати, без да е необходимо да притежават умения за програмиране.**

Според проучване, проведено от Wakefield Research и спонсирано от SAP през 2024 г. [36], в рамките на което са анкетирани 300 висши ръководители на компании с годишни приходи от поне 1 млрд. долара в САЩ: 52% от висшите ръководители се доверяват на ИИ да анализира данни и да дава препоръки за вземане на решения. Други 48% използват ИИ за идентифициране на неочетени преди това рискове, а 47% използват ИИ за предлагане на алтернативни планове. Освен това 40% използват ИИ за разработване на нови продукти, планиране на бюджета и проучване на пазара. Проучването показва и положителното въздействие на ИИ върху личния живот, като 39% от анкетираните съобщават за подобрен баланс между професионалния и личния живот, 38% - за подобрено психично здраве, а 31% - за по-ниски нива на стрес.

При цялата си сила обаче LLM остават инструмент, който е важно да се използва съзнателно. Както всяка технология, те имат ограничения. Един от най-известните проблеми са така наречените "халюцинации" - случаи, при които моделът уверено дава правдоподобен, но фактически неверен отговор. Затова е изключително важно да се разбере как работи моделът: какви данни и модели на данни може да интерпретира без грешка, как интерпретира заявките и откъде получава информацията си. Струва си също така да се помни, че знанията на МЛНЗ са ограничени до датата на неговото обучение и без връзка с външни данни моделът може да не вземе предвид настоящите норми, стандарти, цени или технологии.

Решението на тези проблеми е да се актуализират редовно векторните бази данни, да се свързват със съответните източници и да се разработят автономни агенти с изкуствен интелект - агенти, които не просто отговарят на въпроси, а активно използват данни за обучение, управляват задачи, идентифицират рискове, предлагат варианти за оптимизация и наблюдават изпълнението на проекта.

Преходът към LLM -интерфейси в строителството не е само технологична новост. Това е промяна на парадигмата, която премахва бариерите между хората и данните. Това е възможността да работим с информация толкова лесно, колкото говорим помежду си - и все пак да получаваме точни, проверени и приложими резултати.

Компаниите, които започнат да използват такива инструменти по-рано от останалите, ще получат значително конкурентно предимство. Това включва ускоряване на работата, намаляване на разходите и подобряване на качеството на проектните решения чрез бърз достъп до анализ на данни и възможност за бързо намиране на отговори на сложни въпроси. Но има и въпроси, свързани със сигурността, които трябва да се вземат предвид. Използването на базирани в облака LLM -услуги може да бъде свързано с рискове от изтичане на данни. Затова организациите все по-често търсят алтернативни решения, които им позволяват да внедряват LLM инструменти в собствената си инфраструктура - локално, с пълна защита и контрол върху информацията.

## Използване на местни LLM за чувствителни фирмени данни

Появата на първите чат-LLMs през 2022 г. бележи нов етап в развитието на изкуствения интелект. Веднага след широкото разпространение на тези модели обаче възникна основателен въпрос: доколко е сигурно прехвърлянето на данни и заявки, свързани с компанията, в облака? Повечето базирани в облака езикови модели съхраняват историята на комуникацията и качените документи на своите сървъри, а за компаниите, които се занимават с чувствителна информация, това беше сериозна пречка за внедряването на ИИ.

Едно от най-устойчивите и логични решения на този проблем е разгръщането на отворен код LLM локално, в рамките на корпоративната ИТ инфраструктура. За разлика от облачните услуги, локалните модели работят без интернет връзка, не прехвърлят данни към външни сървъри и дават на компаниите пълен контрол върху информацията

*Най-добрият отворен модел [Open Source LLM] понастоящем е сравним по производителност със затворените модели [като ChatGPT, Claude], но с изоставане от около една година [77].*

- Бен Котие, водещ изследовател в Epoch AI, изследователска организация с нестопанска цел, 2024 г.

Големите технологични компании започнаха да предоставят своите LLM за местно използване. Сериите LLaMA с отворен код на Meta и бързо разрастващият се проект DeepSeek от Китай са примери за преминаване към отворена архитектура. Наред с тях Mistral и Falcon също пуснаха мощни модели, освободени от ограниченията на патентованите платформи. Тези инициативи не само ускориха развитието на глобалния ИИ, но и предоставиха на компаниите, които се грижат за неприкосновеността на личния живот, реални алтернативи за независимост, гъвкавост и съответствие със сигурността.

В корпоративната среда, особено в строителния бранш, защитата на данните не е просто въпрос на удобство, а на спазване на нормативните изисквания. Работата с тръжни документи, оценки, чертежи и поверителна кореспонденция изисква строг контрол. И точно тук местният LLM предлага необходимата гаранция, че данните остават в периметъра на компанията.

	Cloud LLMs (OpenAI, Claude)	Local LLMs (DeepSeek, LLaMA)
<b>Data Control</b>	Data is transmitted to third parties	Data remains within the company's network
<b>License</b>	Proprietary, paid	Open-source (Apache 2.0, MIT)
<b>Infrastructure</b>	Requires internet	Operates in an isolated environment
<b>Customization</b>	Limited	Full adaptation to company needs
<b>Cost</b>	Pay-per-token/request	One-time hardware investment + maintenance costs
<b>Scalability</b>	Easily scalable with cloud resources	Scaling requires additional local hardware
<b>Security &amp; Compliance</b>	Risk of data leaks, may not meet strict regulations (GDPR, HIPAA)	Full compliance with internal security policies
<b>Performance &amp; Latency</b>	Faster inference due to cloud infrastructure	Dependent on local hardware, may have higher latency
<b>Integration</b>	API-based integration, requires internet access	Can be tightly integrated with on-premise systems
<b>Updates &amp; Maintenance</b>	Automatically updated by provider	Requires manual updates and model retraining
<b>Energy Consumption</b>	Energy cost is covered by provider	High power consumption for inference and training
<b>Offline Availability</b>	Not available without an internet connection	Works completely offline
<b>Inference Cost</b>	Pay-per-use model (cost scales with usage)	Fixed cost after initial investment

**Фиг. 3.3-3 Локалните модели осигуряват пълен контрол и сигурност, докато решенията, базирани на облак, предлагат лесна интеграция и автоматични актуализации.**

Основни предимства на местния софтуер с отворен код LLM:

- Пълен контрол върху данните. Цялата информация остава в рамките на компанията, което елиминира неоторизирания достъп и изтичането на данни.
- Офлайн работа. Не зависи от интернет връзка, което е особено важно за изолирани ИТ инфраструктури. Това също така гарантира непрекъсната работа в условията на санкции или блокирани облачни услуги.
- Гъвкавост на приложението. Моделът може да се използва за генериране на текст, анализ на данни, писане на програмен код, подпомагане на проектирането и управление на бизнес процеси.
- Адаптиране към корпоративните цели. LLM може да се обучава по вътрешни документи,

което ви позволява да вземете предвид спецификата на работата на компанията и нейните отраслови особености. Местният LLM може да бъде свързан към CRM, ERP или BI платформи, което ви позволява да автоматизирате анализа на клиентски запитвания, създаването на отчети или дори прогнозирането на тенденции.

Внедряването на безплатния модел с отворен код на DeepSeek -R1-7B на сървър за достъп на цял екип от потребители на цена от 1000 USD на месец може потенциално да струва по-малко от годишните такси за API в облака, като ChatGPT или Claude, и позволява на компаниите да поемат пълен контрол върху своите данни, да елиминират прехвърлянето им в интернет и да спомогнат за спазването на регуляторните изисквания, като GDPR

В други отрасли местните LLM вече променят подхода си към автоматизацията. В областта на услугите за поддръжка те отговарят на чести запитвания на клиенти, като намаляват работното натоварване на операторите. В отделите за човешки ресурси те анализират автобиографии и подбират подходящи кандидати. В електронната търговия те генерират персонализирани оферти, без да разкриват данни за потребителите.

Подобен ефект се очаква и в строителния сектор. Благодарение на интеграцията на LLM с данните и стандартите за проектите е възможно да се ускори подготовката на документацията, да се автоматизира изготвянето на оценки и прогнозни анализи на разходите. Използването на LLM в комбинация със структурирани таблици и рамки за данни се превръща в особено перспективна област.

## **Пълен контрол върху AI в компанията и как да внедрите свой собствен LLM**

Съвременните инструменти позволяват на компаниите да разгърнат локално голям езиков модел (LLM) само за няколко часа. Това дава пълен контрол върху данните и инфраструктурата, като се премахва зависимостта от външни облачни услуги и се свежда до минимум рисъкът от изтичане на информация. Това решение е особено подходящо за организации, работещи с чувствителна проектна документация или чувствителни търговски данни.

В зависимост от задачите и ресурсите се предлагат различни сценарии за внедряване - от готови решения до по-гъвкави и мащабируеми архитектури. Един от най-лесните инструменти е Ollama, който ви позволява да стартирате езикови модели буквально с едно щракване на мишката, без да са ви необходими задълбочени технически познания. Бърз старт с Ollama:

1. Изтеглете дистрибуцията за вашата операционна система (Windows / Linux / macOS) от официалния уебсайт: [ollama.com](https://ollama.com)
2. Инсталрайте модела чрез командния ред. Например, за модела *Mistral*:

Олама тича мистрал

3. След като стартирате модела, той е готов за работа - можете да изпращате текстови заявки през терминала или да го интегрирате в други инструменти. Стартрайте модела и изпълнете заявка:

```
ollama run mistral "Как да съзdam изчисление с всички ресурси за работата по монтирането на преградна стена от гипсокартон с ширина 100 мм?".
```

За тези, които предпочитат да работят в позната визуална среда, има LM Studio - бесплатно приложение с интерфейс, напомнящ ChatGPT

- Инсталирайте LM Studio, като изтеглите комплекта за разпространение от официалния уебсайт - lmstudio.ai
- Чрез вградения каталог изберете модел (напр. Falcon или GPT-Neo-X) и го изтеглете
- Работете с модела чрез интуитивен интерфейс, наподобяващ ChatGPT, но напълно локализиран.

	Developer	Parameters	GPU Requirements (GB)	Features	Best For
Mistral 7B	Mistral AI	7	8 (FP16)	Fast, supports multimodal tasks (text + images), fully open-source code	Lightweight tasks, mobile devices, laptops
LLaMA 2	Meta	7–70	16–48 (FP16)	High text generation accuracy, adaptable for technical tasks, CC-BY-SA license	Complex analytical and technical tasks
Baichuan 7B/13B	Baichuan Intelligence	7–13	8–16 (FP16)	Fast and efficient, great for large data processing, fully open-source code	Data processing, automating routine tasks
Falcon 7B/40B	Technology Innovation Institute (TII)	7–40	8–32 (FP16)	Open-source, high performance, optimized for fast work	Workloads with limited computational resources
DeepSeek-V3	DeepSeek	671	1543 (FP16) / 386 (4-bit)	Multilingual, 128K token context window, balanced speed and accuracy	Large enterprises, SaaS platforms, multitasking scenarios
DeepSeek-R1-7B	DeepSeek	7	18 (FP16) / 4.5 (4-bit)	Retains 92% of R1 capabilities in MATH-500, local deployment support	Budget solutions, IoT devices, edge computing

Фигура 3.3-4 Сравнение на популярни местни LLM модели с отворен код.

Изборът на модел зависи от изискванията за скорост, точност и наличните хардуерни възможности (Фигура 3.3-4). Малките модели, като Mistral 7B и Baichuan 7B, са подходящи за леки задачи и

мобилни устройства, докато мощните модели, като DeepSeek -V3, изискват значителни изчислителни ресурси, но предлагат висока производителност и поддръжка на множество езици. През следващите години пазарът на LLM ще се разраства бързо - ще виждаме все повече леки и специализирани модели. Вместо LLM с общо предназначение, които обхващат цялото човешко съдържание, ще се появят модели, обучени на тесен експертен опит в дадена област. Например, можем да очакваме появата на модели, предназначени единствено за обработка на инженерни изчисления, строителни оценки или данни в CAD формати. Такива специализирани модели ще бъдат по-бързи, по-точни и по-безопасни за използване - особено в професионални среди, където високата надеждност и задълбочеността на темата са важни.

След стартирането на местния LLM той може да бъде адаптиран към специфичните задачи на компанията. За тази цел се използва техниката за фина настройка, при която моделът се дообучава върху вътрешни документи, технически инструкции, образци на договори или проектна документация.

## RAG: Интелигентен LLM - помощници с достъп до корпоративни данни

Следващият етап в развитието на приложението на LLM в бизнеса е интегрирането на моделите с реални корпоративни данни в реално време. Този подход се нарича RAG (Retrieval-Augmented Generation) - извличане на данни и разширено генериране. В тази архитектура езиковият модел се превръща не просто в диалогов интерфейс, а в пълноценен интелигентен асистент, способен да навигира в документи, чертежи, бази данни и да предоставя точни, контекстуални отговори.

Основното предимство на RAG е възможността да се използват вътрешни данни на компанията, без да е необходимо предварителното обучение на модела, като същевременно се поддържа висока точност и гъвкавост при обработката на информацията.

Технологията RAG съчетава два основни компонента:

- **Извличане:** моделът се свързва с хранилища на данни - документи, таблици, PDF - файлове, чертежи - и извлича съответната информация, както е поискано от потребителя.
- **Разширено генериране:** въз основа на извлечените данни моделът генерира точен и информиран отговор, като взема предвид контекста и спецификата на заявката.

За да стартирате LLM с подкрепата на RAG, трябва да следвате няколко стъпки:

- **Подготовка на данните:** събиране на необходимите документи, чертежи, спецификации, таблици. Те могат да бъдат в различни формати и структури - от PDF до Excel.
- **Индексиране и векторизиране:** с помощта на инструменти като LlamaIndex или LangChain данните се преобразуват във векторни представления, които ви позволяват да намирате семантични връзки между текстови фрагменти (повече за векторните бази данни и преобразуването на големи масиви във векторни представления, включително проекти на CAD, в Част 8).
- **Запитване към асистента:** след като данните бъдат качени, можете да зададете въпроси

на модела и той ще търси отговори в корпоративната рамка, а не в общите знания, събрани от интернет.

Да предположим, че дадена компания има папка constructionsite\_docs, в която се съхраняват договори, инструкции, оценки и таблици. С помощта на скрипт на Python (фиг. 3.3-5) можем да сканираме тази папка и да изградим векторно индексиране: всеки документ ще бъде преобразуван в набор от вектори, отразяващи семантичното съдържание на текста. По този начин документите се превръщат в своеобразна "карта на значенията", по която моделът може ефективно да навигира и да намира връзки между термини и фрази.

Например, моделът "помни", че думите "връщане" и "рекламация" често се срещат в раздела на договора, отнасящ се до транспортирането на материали до строителната площадка. Тогава, ако бъде зададен въпрос - например "Какъв е срокът за връщане на стоката?", той може да отговори. (фиг. 3.3-5 - ред 11 от кода) - LLM ще анализира вътрешните документи и ще намери точната информация, действайки като интелигентен асистент, способен да чете и разбира съдържанието на всички корпоративни файлове.

```

1 from llama_index import SimpleDirectoryReader, VectorStoreIndex
2
3 # Load documents from the folder
4 documents = SimpleDirectoryReader("constructionsite_docs").load_data()
5
6 # Creating a vector index for semantic search
7 index = VectorStoreIndex.from_documents(documents)
8
9 # Integration with LLM (e.g. Llama 3)
10 query_engine = index.as_query_engine()
11 response = query_engine.query("What are the return terms in the contracts?")
12 print(response)

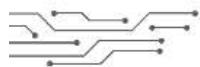
```

**Фиг. 3.3-5 LM чете папка с файлове - подобно на начина, по който човек я отваря и търси необходимия документ**

Кодът може да се изпълнява на всеки компютър с инсталиран Python. Ще говорим повече за използването на Python и IDE за изпълнение на кода в следващата глава.

Местното внедряване на LLM не е просто тенденция, а стратегическо решение за компании, които ценят сигурността и гъвкавостта. Внедряването на LLM, независимо дали на локални фирмени компютри или чрез онлайн решения, обаче е само първата стъпка. За да приложат възможностите на LLM към реални задачи, компаниите трябва да използват инструменти, които им позволяват не само да получават отговори от чата, но и да съхраняват създадената логика под формата на код, който може да се изпълнява извън контекста на използване на LLM. Това е важно за мащабирането на решенията - правилно организираните процеси позволяват разработките на ИИ да се прилагат едновременно към няколко проекта или дори в цялата компания.

В този контекст изборът на подходяща среда за разработка (IDE) играе важна роля. Съвременниятите инструменти за програмиране позволяват не само да се разработват решения, базирани на LLM, но и да се интегрират в съществуващи бизнес процеси, превръщайки ги в автоматизирани ETL -Пилоти



## ГЛАВА 3.4.

### IDE С ПОДДРЪЖКА НА LLM И БЪДЕЩИ ПРОГРАМНИ ПРОМЕНИ

#### Избор на IDE: от експерименти по LLM до бизнес решения

Когато навлизате в света на автоматизацията, анализа на данни и изкуствения интелект - особено когато работите с големи езикови модели (LLM) - е изключително важно да изберете правилната интегрирана среда за разработка (IDE). Тази IDE ще бъде вашият основен работен инструмент: мястото, където ще се изпълнява кодът, генериран от LLM, както на локален компютър, така и в рамките на корпоративната мрежа. Изборът на IDE определя не само удобството на работата ви, но и колко бързо ще можете да преминете от експериментални LLM заявки към пълноценни решения, вградени в реални бизнес процеси.

**IDE** (интегрирана среда за разработка) е универсален градивен елемент на вашия компютър за автоматизация на процеси и обработка на данни. Вместо да държите по-отделно трион, чук, бормашина и други инструменти, разполагате с едно устройство, което може да прави всичко - да реже, закрепва, пробива и дори да проверява качеството на материалите. IDE за програмисти е единно пространство, в което можете да пишете код (по аналогия със строителството - да създавате чертежи), да тествате работата му (сглобяване на модел на сграда), да откривате грешки (като проверка на здравината на конструкциите в строителството) и да стартирате готовия проект (пускане на къщата в експлоатация).

Преглед на популярните IDE:

- **PyCharm®** (JetBrains) е мощна професионална среда за разработка на Python. Тя е подходяща за сериозни проекти поради големия брой вградени функции. Основната поддръжка на интерактивни файлове Jupyter (IPYNB) обаче е налична само в платената версия, а за начинаещите интерфейсът може да се окаже непосилен.

Файлът с разширение IPYNB (Interactive Python Notebook) е формат за интерактивни бележници Jupyter® (фигура 3.4-1), в който кодът, визуализациите и обясненията са обединени в един документ. Този формат е идеален за създаване на отчети, анализи и сценарии за обучение.

- **VS Code®** (Microsoft) е бърз, гъвкав и адаптивен инструмент с безплатна поддръжка на IPYNB и много приставки. Подходящ е както за начинаещи, така и за професионалисти. Позволява интегрирането на GitHub Copilot и плъгини за езикови модели, което го прави чудесен избор за проекти за изкуствен интелект и наука за данните.
- **Jupyter Notebook** - класически и популярен избор за експериментиране и учене. Тя ви позволява да пишете код, да добавяте обяснения и да визуализирате резултатите в един интерфейс (фиг. 3.4-1). Идеален е за бързо тестване на хипотези, работа с LLM и създаване на възпроизвеждани диви стъпки за анализ на данни. За управление на зависимостите и библиотеките се препоръчват специални инструменти като pip и conda.

лиотеките препоръчваме да използвате Anaconda Navigator - визуален интерфейс за управление на средата на Python.



**Фиг. 3.4-1 Jupyter Notebook един от най-удобните и популярни инструменти за създаване на Pipeline процеси.**

- **Google Collab™** (и платформата Kaggle (Фигура 9.2-5)) е облачна алтернатива на Jupyter, която осигурява безплатен достъп до GPU/TPU. Това е чудесно решение за започване на работа - без инсталација на локален софтуер и с възможност за работа директно от браузър. Поддържа интеграция с Google Drive, а от скоро и с Gemini (LLM на Google).

	PyCharm	VS Code	Jupyter Notebook	Google Colab
<b>Complexity</b>	High	Medium	Low	Low
.ipynb support	Paid	Free	Built-in	Built-in
Copilots	Yes	Yes	Yes	Yes
Computing resources	Local	Local	Local	Cloud
For whom	Professionals	Universal	Beginners	Experimenters

**Фигура 3.4-2 Сравнение на IDE: Jupyter Notebook един от най-удобните и лесни инструменти за създаване на Pipeline процеси.**

Изборът на IDE зависи от задачите ви. Ако искате бързо да започнете работа с изкуствен интелект,

опитайте Jupyter Notebook или Google Collab. За сериозни проекти е по-добре да използвате PyCharm или VS Code. Основното нещо е да започнете работа. Съвременните инструменти ви позволяват бързо да превърнете експериментите си в работещи решения.

Всички IDE, описани по-горе, ви позволяват да създавате конвейери за обработка на данни - т.е. вериги от блокови модули за код (които могат да бъдат генериирани от LLM), всеки от които отговаря за различен етап, например:

- аналитични сценарии,
- вериги за извлечане на информация от документи,
- автоматични отговори въз основа на RAG,
- генериране на отчети и визуализации.

Благодарение на модулната структура всяка стъпка може да бъде представена като отделен блок: зареждане на данни → филтриране → анализ → визуализация → експортиране на резултатите. Тези блокове могат да се използват многократно, - да се адаптират и сглобяват в нови вериги като конструктор, само за данни.

За инженерите, мениджърите и анализаторите това открива възможност за документиране на логиката на вземане на решения под формата на код, който може да бъде генериран с LLM. Този подход спомага за ускоряване на рутинни задачи, автоматизиране на типични операции и създаване на повторяеми процеси, при които всяка стъпка е ясно документирана и прозрачна за всички членове на екипа.

Автоматизираните ETL Pipelines (фиг. 7.2-3), Apache Airflow (фиг. 7.4-4), Apache NiFi (фиг. 7.4-5) и n8n (фиг. 7.4-б) инструменти за изграждане на логически блокове за автоматизация на процесите ще бъдат разгледани по-подробно в част 7 и част 8 на книгата.

## **IDE с поддръжка на LLM и бъдещи промени в програмирането**

Интегрирането на изкуствения интелект в процесите на разработка променя пейзажа на програмирането. Съвременните среди вече не са просто текстови редактори с подчертаване на синтаксиса - те се превръщат в интелигентни асистенти, способни да разбират логиката на проекта, да допълват кода и дори да обясняват как работи определен фрагмент от кода. На пазара се появяват продукти, които използват изкуствен интелект, за да разширят границите на конвенционалната разработка:

- **GitHub Copilot** (интегрира се с VS Code, PyCharm): AI-асистент, който генерира код въз основа на коментари или частични описание, като превръща текстовите подсказки в готови решения.
- **Cursor** (разклонение на VS Code с AI-kernel): позволява не само да завършите писането на код, но и да задавате въпроси към проекта, да търсите зависимости и да се учате от базата с код.
- **JetBrains AI Assistant**: приставка за JetBrains IDE (включително PyCharm) с функция за обяснение на сложен код, оптимизация и създаване на тестове.
- **Amazon CodeWhisperer**: аналог на Copilot с акцент върху сигурността и поддръжката на

услугите на Amazon AWS.

Програмирането ще претърпи драматична промяна през следващите години. Основният фокус ще се измести от рутинното писане на код към проектирането на модели и архитектура на данните - разработчиците ще бъдат по-ангажирани с проектирането на системата, докато AI ще поеме шаблонни задачи: генериране на код, тестове, документация и основни функции. Бъдещето на програмирането е сътрудничество между хората и ИИ, при което машините поемат техническата рутина, а хората се фокусират върху творчеството.

Програмирането на естествен език ще стане ежедневие. Персонализирането на IDE ще достигне ново ниво - средите за разработка ще се научат да се адаптират към стила на работа на потребителя и неговите компании, като предвиждат модели, предлагат контекстуални решения и се учат от предишни проекти

Това не отменя ролята на разработчика, но ще я промени радикално: от писане на код към управление на знанията, качеството и процесите. Тази еволюция ще засегне и бизнес разузнаването, където създаването на отчети, визуализации и приложения за подпомагане на вземането на решения все повече ще се извършва чрез генериране на код и логика с помощта на ИИ и LLM, чат и агентски интерфейси.

След като компанията е създала LLM чатове и е избрала подходяща среда за разработка, следващата важна стъпка е да организира данните. Този процес включва извлечане на информация от различни източници, нейното почистване, трансформиране в структурирана форма и интегрирането ѝ в корпоративните системи.

При съвременния подход към управлението на данни, ориентиран към данните, основната цел е данните да бъдат приведени в единна универсална форма, която да е съвместима с голям брой инструменти и приложения. Необходими са специализирани библиотеки, които да се справят с процесите на структуриране и структурираните данни. Една от най-мощните, гъвкави и популярни е библиотеката Pandas за Python. Тя ви позволява удобно да обработвате таблични данни: филтриране, групиране, почистване, добавяне, агрегиране и изготвяне на отчети.

## Python Pandas: незаменим инструмент за работа с данни

Pandas заема специално място в света на анализа и автоматизацията на данни. Това е една от най-популярните и широко използвани библиотеки на езика за програмиране Python [86], предназначена за работа със структурирани данни.

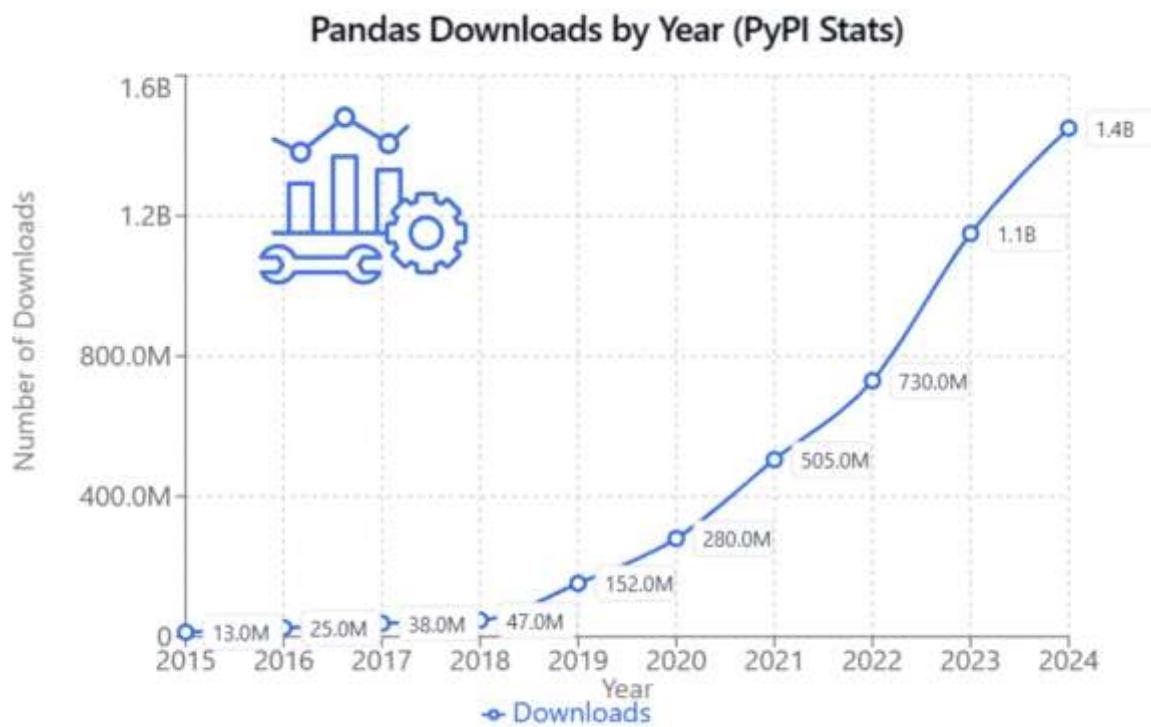
Библиотеката е като набор от готови инструменти: функции, модули, класове. Точно както на строителната площадка не е необходимо всеки път да измисляте чук или ниво, така и в програмирането библиотеките ви позволяват бързо да решавате проблеми, без да изобретявате наново основните функции и решения.

Pandas е библиотека с отворен код за Python, която предоставя високопроизводителни и интуитивни структури от данни, по-специално DataFrame, универсален формат за работа с таблици. Pandas е швейцарско ножче за анализатори, инженери и разработчици, работещи с данни.

Python е език за програмиране от високо ниво с прост синтаксис, който се използва активно в областта на анализа, автоматизацията, машинното обучение и уеб разработката. Неговата популярност се дължи на четимостта на кода, междуплатформения му характер и богатата екосистема от библиотеки. Към днешна дата за Python са създадени повече от 137 000 пакета с отворен код [87], като този брой продължава да нараства почти ежедневно. Всяка такава библиотека е своеобразно хранилище на готови функции: от прости математически операции до сложни инструменти за обработка на изображения, анализ на големи обеми от данни, невронни мрежи и интеграция с външни услуги.

С други думи, представете си, че имате свободен и отворен достъп до стотици хиляди готови софтуерни решения - библиотеки и инструменти, които можете директно да вградите в бизнес процесите си. Това е като огромен каталог от приложения за автоматизация, анализ, визуализация, интеграция и други - и всички те са достъпни веднага след инсталациейта на Python.

Pandas е един от най-受欢迎ните пакети в екосистемата на Python. През 2022 г. средният брой на изтеглянията на библиотеката Pandas достига 4 милиона на ден (Фигура 3.4-3), докато в началото на 2025 г. тази цифра се увеличава до 12 милиона изтегляния на ден, което отразява нарастващата ѝ популярност и широкото ѝ използване в анализите на данни и чатовете на LLM [86].



Фигура 3.4-3 Pandas е една от най-изтегляните библиотеки. През 2024 г. годишният брой на нейните изтегляния надхвърля 1,4 милиарда.

Езикът за заявки в библиотеката Pandas е сходен по функционалност с езика за заявки SQL, който разглеждахме в главата "Релационни бази данни и език за заявки SQL".

В света на анализите и управлението на структурирани данни Pandas се отличава със своята простота, бързина и мощ, като предоставя на потребителите широк набор от инструменти за ефективен анализ и обработка на информация.

И двата инструмента - SQL и Pandas - предоставят мощни възможности за манипулиране на данни, особено в сравнение с традиционния Excel. Те поддържат операции като селектиране, филтриране (фигура 3.4-4), като единствената разлика е, че SQL е оптимизиран за работа с релационни бази данни, докато Pandas обработва данни в RAM, което му позволява да работи на всеки компютър, без да е необходимо да се създават бази данни и да се внедрява отделна инфраструктура.

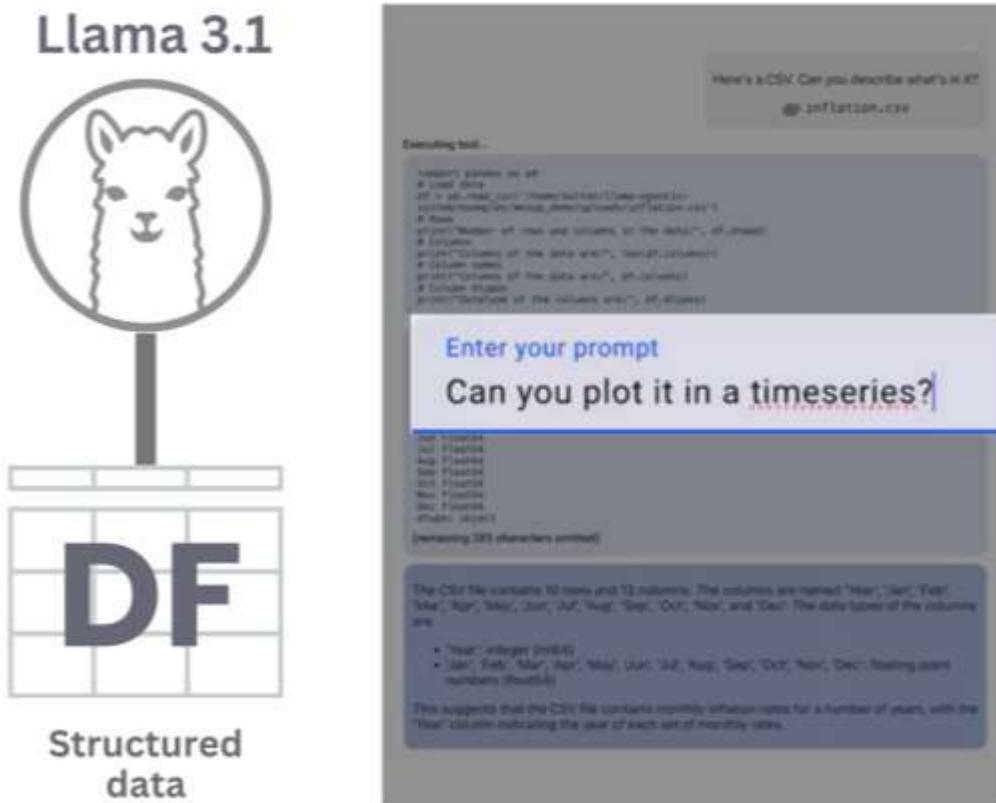


Фиг. 3.4-4 Пандас, за разлика от SQL, има гъвкавостта да работи с различни формати данни, които не се ограничават до бази данни.

Pandas често се предпочита за научни изследвания, автоматизация на процеси, създаване на конвейери (включително ETL) и манипулиране на данни в Python, докато SQL е стандарт за управление на бази данни и често се използва в корпоративни среди за обработка на големи количества данни.

Библиотеката Pandas на езика за програмиране Python ви позволява да извършвате не само основни операции, като четене и писане на таблици, но и по-сложни задачи, включително обединяване на данни, групиране на данни и извършване на сложни аналитични изчисления.

Днес библиотеката Pandas се използва не само в академичните изследвания и бизнес анализите, но и във връзка с LLM -моделите. Например подразделението Meta® (Facebook™) при публикуването на новия модел с отворен код LLaMA 3.1 през 2024 г. обръща специално внимание на работата със структурирани данни, като един от ключовите и първи случаи в неговото издание е именно обработката на структурирани рамки от данни (фиг. 3.4-5) във формат CSV и интеграцията с библиотеката Pandas директно в чата.



Фиг. 3.4-5 Един от първите и основни казуси на екипа на Meta, представен в LlaMa 3.1 през 2024 г., беше изграждането на приложения с помощта на Pandas.

*Pandas е основен инструмент за милиони учени, които обработват и подготвят данни за генеративен изкуствен интелект. Ускоряването на Pandas с нулеви промени в кода ще бъде огромна стъпка напред. Учените, занимаващи се с данни, ще могат да обработват данни за минути вместо за часове и да получават порядъчно повече данни за обучение на генеративни модели на ИИ [88].*

- Дженсен Хуанг, основател и главен изпълнителен директор на NVIDIA

С помощта на Pandas, е възможно да се управляват и анализират набори от данни, които далеч надхвърлят възможностите на Excel. Докато Excel обикновено може да обработва до 1 милион реда данни, Pandas може лесно да обработва набори от данни (Фигура 9.1-2, Фигура 9.1-10), съдържащи десетки милиони редове [89]. Тази способност позволява на потребителите да извършват сложен анализ и визуализация на данни върху големи масиви от данни, като осигуряват задълбочени прозрения и улесняват вземането на решения, основани на данни. Освен това Pandas има силна подкрепа от общността [90]: стотици милиони разработчици и анализатори по света (Kaggle.com, Google Collab, Microsoft® Azure™ Notebooks, Amazon SageMaker) го използват онлайн или офлайн всеки ден, осигурявайки голям брой готови решения за всеки бизнес проблем.

В основата на повечето аналитични процеси в Python е структурирана форма на данни, наречена DataFrame, предоставена от библиотеката Pandas. Тя е мощен и гъвкав инструмент за организиране, анализ и визуализиране на таблични данни.

## DataFrame: универсален формат за таблични данни

DataFrame е централната структура в библиотеката Pandas, която представлява двуизмерна таблица (Фигура 3.4-6), в която редовете съответстват на отделни обекти или записи, а колоните - на техните характеристики, параметри или категории. Тази структура визуално наподобява електронните таблици на Excel, но е много по-добра по отношение на гъвкавост, мащабируемост и функционалност.

**DataFrame** е начин за представяне и обработка на таблични данни, съхранявани в оперативната памет на компютъра.

DataFrame е начин за представяне и обработка на таблични данни, съхранявани в оперативната памет на компютъра. В една таблица редовете могат да отразяват например елементи на строителен проект, а колоните - техните свойства: категории, размери, координати, разходи, термини и т.н. Освен това такава таблица може да съдържа както информация за един проект (фиг. 4.1-13), така и данни за милиони обекти от хиляди различни проекти (фиг. 9.1-10). Благодарение на векторизираните операции на Pandas е лесно да се филтрират, групират и агрегират такива обеми от информация с висока скорост.

The diagram illustrates a construction project as a structured data frame. On the left, there is a small illustration of a building under construction with a crane. On the right, there is a large table with the following data:

	ID	Name	Category	Family Name	Height	BoundingBoxMin_X	BoundingBoxMin_Y	BoundingBoxMin_Z	Level
431144	Single-Flush	OST_Doors	Single-Flush	6.88976378	20.1503	-10.438	9.84252		Level 1
431198	Single-Flush	OST_Doors		6.88976378	13.2281	-1.1207	9.84252		Level 2
457479	Single Window	OST_Windows	Single Window	8.858267717	-11.434	-11.985	9.80971		Level 2
485432	Single Window	OST_Windows	Single Window	8.858267717	-11.434	4.25986	9.80971		Level 2
490150	Single-Flush	OST_Doors	Single-Flush	6.88976378	-1.5748	-2.9565	-1E-16		Level 1
493697	Basic Wall	OST_Walls	Basic Wall		-38.15	20.1656	-4.9213		Level 1
497540	Basic Wall	OST_Walls	Basic Wall		-4.5212	0.0708	9.84252		Level 1

Фигура 3.4-6 Строителен проект като рамка от данни е двуизмерна таблица с елементи в редове и атрибути в колони.

Nvidia изчислява, че днес вече до 30% от всички компютърни ресурси се използват за обработка на структурирани данни - рамки от данни - и този дял продължава да расте.

*Обработката на данни е това, което вероятно една трета от компютърните дейности в света се извършва във всяка компания. Обработката на данни и данните на повечето компании са в DataFrame, във формат на таблица*

- Дженсън Хуанг, главен изпълнителен директор на Nvidia [91]

Нека изброим някои основни характеристики на DataFrame в Pandas:

- **Колони:** в DataFrame данните са организирани в колони, всяка от които има уникално име. Атриутните колони могат да съдържат данни от различни типове, подобно на колоните в базите данни или колоните в таблиците.
- **Pandas Серията** е едноизмерна структура от данни в Pandas, подобна на списък или колона в таблица, където всяка стойност съответства на различен индекс. Серията Pandas има над 400 атрибута и метода, което прави работата с данни изключително гъвкава. Можете директно да приложите една от четиристотинте налични функции към дадена колона, да извършвате математически операции, да филтрирате данни, да заменяте стойности, да работите с дати, низове и много други. Освен това Series поддържа векторизирани операции, което значително ускорява обработката на големи масиви от данни в сравнение с цикличните изчисления. Например можете лесно да умножавате всички стойности по число, да замествате липсващи данни или да прилагате сложни трансформации, без да пишете сложни цикли.
- **Редове:** в DataFrame могат да се индексират с уникални стойности. Този индекс ви позволява бързо да променяте и коригирате данните в определени редове.
- **Индекс:** По подразбиране, когато създавате DataFrame Pandas, на всеки ред се присвоява индекс от 0 до N-1 (където N е броят на всички редове в DataFrame). Индексът обаче може да бъде променен, за да включва специални обозначения като дати или уникални характеристики.
- **Индексиране на редове** в DataFrame означава, че на всеки ред се присвоява уникално име или етикет, който се нарича индекс на DataFrame.
- **Типове данни:** DataFrame поддържа различни типове данни, включително: `int`, `float`, `bool`, `datetime64` и `object` за текстови данни. Всяка колона на DataFrame има свой собствен тип данни, който определя какви операции могат да се извършват върху нейното съдържание.
- **Операции с данни:** DataFrame поддържа широк набор от операции за обработка на данни, включително агрегиране (`groupby`), обединяване (`merge` и `join`), конкатенация (`concat`), разделяне-прилагане-съединяване и много други техники за преобразуване на данни.
- **Манипулиране на размера:** DataFrame ви позволява да добавяте и премахвате колони и редове, което го прави динамична структура, която може да се променя според нуждите ви за анализ на данни.
- **Визуализиране на данни:** използвайки вградени техники за визуализация или взаимодействие с популярни библиотеки за визуализация на данни, като Matplotlib или Seaborn,

DataFrame може лесно да се преобразува в графики и диаграми за графично представяне на данните.

- **Въвеждане и извеждане на данни:** Pandas предоставя функции за четене, импортиране и експортиране на данни в различни файлови формати, като CSV, Excel, JSON, HTML и SQL, което потенциално може да превърне DataFrame в централен център за събиране и разпространение на данни.

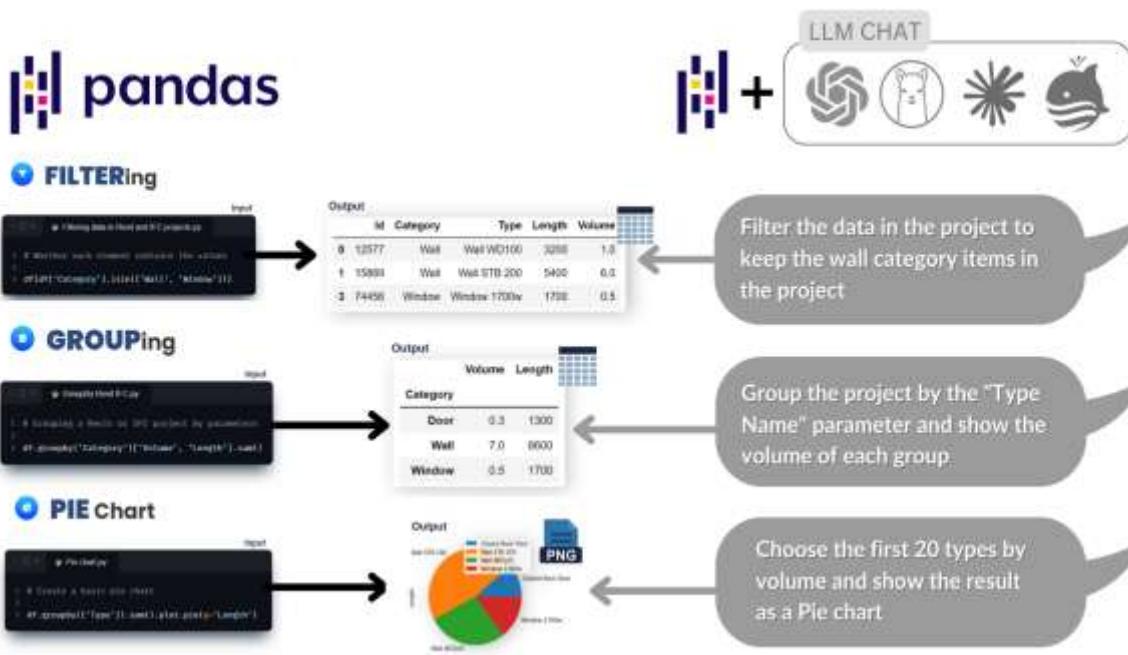
За разлика от CSV и XLSX, Pandas DataFrame осигурява по-голяма гъвкавост и производителност при работа с данни: може да обработва големи количества информация в оперативната памет, поддържа разширени типове данни (включително дати, логически стойности и времеви редове) и предоставя широки възможности за филтриране, агрегиране, обединяване и визуализиране на данни. Докато CSV не съхранява информация за типовете и структурата на данните, а XLSX често е претоварен с форматиране и има ниска мащабируемост, DataFrame остава оптималният избор за бърз анализ, автоматизация на процесите и интеграция с AI -модели (фигура 3.4-7). В следващите глави всеки от тези аспекти на данните ще бъде разгледан подробно, също така в осмата част на книгата ще бъдат разгледани подробно подобни формати като Parquet, Apache Orc, JSON, Feather, HDF5 и складове за данни (фиг. 8.1-2).

	XLSX	CSV	Pandas DataFrame
Storage	Tabular	Tabular	Tabular
Usage	Office tasks, data presentation	Simple data exchange	Data analysis, manipulation
Compression	Built-in	None	None (in-memory)
Performance	Low	Medium	High (memory dependent)
Complexity	High (formatting, styles)	Low	Low
Data Type Support	Limited	Very limited	Extended
Scalability	Low	Low	Medium (memory limited)



Фиг. 3.4-7 DataFrame е оптималният избор за манипулиране на данни с висока производителност и разширена поддръжка на типове данни.

Благодарение на гъвкавостта, мощта и лекотата на използване библиотеката Pandas и форматът DataFrame се превърнаха във фактически стандарт за анализ на данни в Python. Те са идеални както за създаване на прости отчети, така и за изграждане на сложни аналитични конвейери, особено в съчетание с LLM модели.



Фиг. 3.4-8 LLM опростява взаимодействието с Pandas: вместо код е достатъчна текстова заявка.

Днес Pandas се използва активно в базирани на LLM чатове, като ChatGPT, LLaMa, DeepSeek, QWEN и други. В много случаи, когато даден модел получи заявка, свързана с обработка на таблици, валидиране на данни или анализ, той генерира код именно с помощта на библиотеката Pandas. Това превръща DataFrame в естествен "език" за представяне на данни в диалози с изкуствен интелект (Фигура 3.4-8).

Съвременните технологии за данни, като например Pandas, улесняват анализа, автоматизацията и интегрирането на данните в бизнес процесите. Те дават бързи резултати, намаляват натоварването на специалистите и осигуряват повторяемост на операциите.

## Следващи стъпки: изграждане на устойчива рамка за данни

В тази част разглеждахме основните видове данни, използвани в строителната индустрия, запознахме се с различните формати за тяхното съхранение и анализирахме ролята на съвременните инструменти, включително LLM и IDE, при обработката на информацията. Научихме, че ефективното управление на данните е в основата на вземането на информирани решения и автоматизацията на бизнес процесите. Организациите, които са в състояние да структурират и организират своите данни, получават значително конкурентно предимство на етапите на обработка и трансформация на данните.

За да обобщим тази част, си струва да подчертаем основните практически стъпки, които ще ви помогнат да прилагате обсъжданите подходи в ежедневните си задачи:

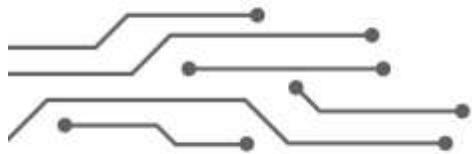
- Извършване на одит на данните за вашите процеси

- Направете опис на всички типове данни, които използвате в проектите си
- Определете кои типове данни и модели са най-важни за вашите бизнес процеси
- Идентифициране на проблемни области, в които информацията често остава неструктуррирана, лошо структурирана или недостъпна
- Започнете да формирате стратегия за управление на данни**
- Повдигане на въпроси, свързани с политиката и стандартите за работа с различни видове данни
- Анализирайте кои от вашите работни процеси могат да бъдат подобрени чрез преобразуване на неструктурирани данни в структурирани.
- Създаване на политика за съхранение на данни и достъп до тях, която отчита сигурността и поверителността.
- Инсталиране и усвояване на основни инструменти за работа с данни**
- Изберете подходяща среда за разработка, която да отговаря на задачите ви (напри мер инсталирайте VS Code или Jupyter Notebook)
- Опитайте се да създадете местен LLM, който да обработва поверително личните ви данни.
- Започнете да експериментирате с библиотеката Pandas, за да обработвате таблични данни XLSX
- Опишете на LLM типичните задачи, с които се справяте в инструменти за електронни таблици или бази данни, и поискайте от LLM да автоматизира работата с помощта на Pandas

Прилагането на тези стъпки ще ви позволи постепенно да трансформирате подхода си към работата с данни, като преминете от разнородни, неструктурирани масиви от данни към единна екосистема, в която данните се превръщат в достъпен и разбираем актив. Започнете с малки стъпки - създайте първата си DataFrame в Pandas, стартирайте локален LLM, автоматизирайте първата си рутинна задача с помощта на Python (напр. електронни таблици в Excel).

Четвъртата част на книгата ще бъде посветена на качеството на данните, организацията, структурирането и моделирането на данните. Ще се спрем на методологии, които превръщат разнородни източници на данни - от PDF файлове и текстове до изображения и CAD модели - в структурирани набори от данни, подходящи за анализ и автоматизация. Ще проучим също така как се формализират изискванията към данните, как се изграждат концептуални и логически модели в строителните проекти и как съвременните езикови модели (LLM) могат да помогнат в този процес.





## IV ЧАСТ

### **КАЧЕСТВО НА ДАННИТЕ: ОРГАНИЗАЦИЯ, СТРУКТУРИРАНЕ, МОДЕЛИРАНЕ**

Четвъртата част е посветена на методологиите и технологиите за преобразуване на разнородна информация в структурирани масиви от данни с високо качество. Подробно са разгледани процесите на формиране и документиране на изискванията към данните като основа за ефективна информационна архитектура в строителните проекти. Представени са практически методи за извлечане на структурирана информация от различни източници (PDF -документи, изображения, текстови файлове, CAD -модели) с примери за прилагане. Анализирано е използването на регулярни изрази (RegEx) и други инструменти за автоматично валидиране и проверка на данни. Процесът на моделиране на данни на концептуално, логическо и физическо ниво е описан стъпка по стъпка, като се отчита спецификата на строителната индустрия. Демонстрирани са конкретни примери за използване на езикови модели (LLM) за автоматизиране на процесите на структуриране и валидиране на информацията. Предложени са ефективни подходи за визуализация на резултатите от анализа, което увеличава достъпността на аналитичната информация за всички нива на управление на строителните проекти.

## ГЛАВА 4.1.

### ПРЕОБРАЗУВАНЕ НА ДАННИ В СТРУКТУРИРАНА ФОРМА

В ерата на икономиката, основана на данни, данните се превръщат в основа за вземане на решения, а не в пречка. Вместо постоянно да адаптират информацията към всяка нова система и нейните формати, компаниите все повече се стремят да формират единен структуриран модел на данните, който да служи като универсален източник на истина за всички процеси. Съвременните информационни системи се проектират не около форматите и интерфейсите, а около смисъла на данните - защото структурата може да се промени, но смисълът на информацията остава един и същ за много по-дълго време.

Ключът към ефективната работа с данни не е в безкрайното им преобразуване и трансформиране, а в правилното им организиране от самото начало: създаване на универсална структура, способна да осигури прозрачност, автоматизация и интеграция на всички етапи от жизнения цикъл на проекта.

Традиционният подход налага ръчни корекции при всяко внедряване на нова платформа: миграция на данни, промяна на имената на атрибутите, коригиране на форматите. Тези стъпки не подобряват качеството на самите данни, а само прикриват проблемите, създавайки порочен кръг от безкрайни трансформации. В резултат на това компаниите стават зависими от конкретни софтуерни решения, а цифровата трансформация се забавя.

В следващите глави ще разгледаме как правилно да структурираме данните, а след това как да създадем универсални модели, да сведем до минимум зависимостта от платформата и да се съсредоточим върху най-важното - данните като стратегически ресурс, около който се изграждат устойчиви процеси.

#### **Научете се как да превръщате документи, PDF, снимки и текстове в структурирани формати**

В строителните проекти по-голямата част от информацията съществува в неструктурен вид: технически документи, работни задания, чертежи, спецификации, графици, протоколи. Тяхното разнообразие - както по формат, така и по съдържание - затруднява интеграцията и автоматизацията.

Процесът на преобразуване в структурирани или полуструктурни формати може да варира в зависимост от вида на входните данни и желаните резултати от обработката.

Преобразуването на данните от неструктурена в структурирана форма е едновременно изкуство и наука. Този процес варира в зависимост от вида на входните данни и целта на анализа и често заема значителна част от работата на инженера по данните (фигура 3.2-5) и на анализатора, като целта е да се получи чист, организиран набор от данни.



**Фиг. 4.1-1 Конвертиране на неструктурен сканиран документ в структуриран табличен формат.**

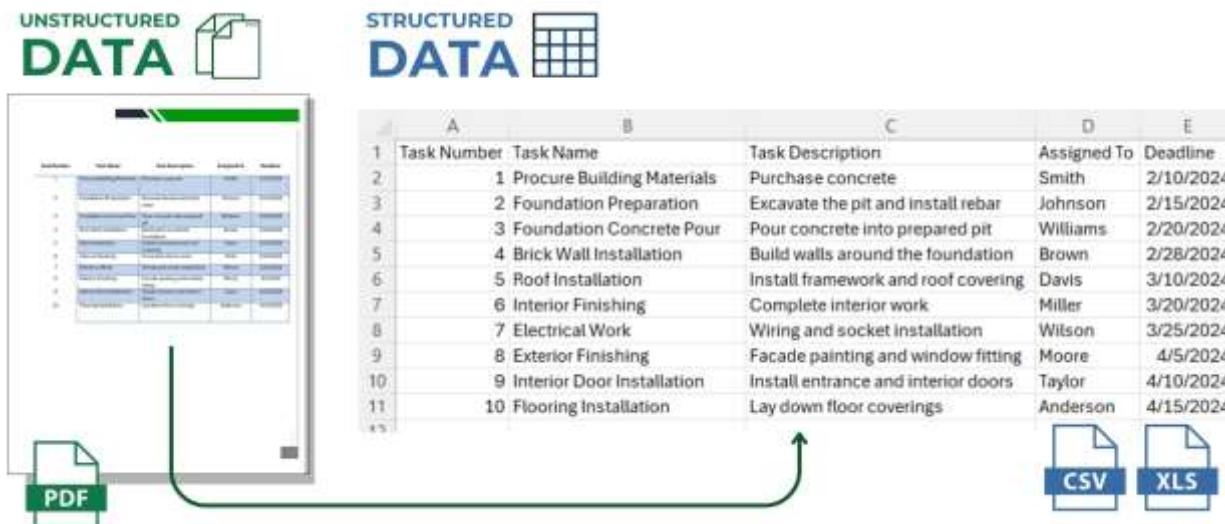
Превръщането на документи, PDF, снимки и текстове в структуриран формат (фигура 4.1-1) е процес, който включва следните стъпки: стъпка по стъпка:

- **Извличане**): На тази стъпка се зарежда документ или изображение, съдържащо неструктурни данни. Това може да бъде например PDF -документ, снимка, чертеж или схема.
- **Трансформиране на данни (Transform)**: Следва стъпка на трансформиране на неструктурни данни в структуриран формат. Например това може да включва разпознаване и интерпретиране на текст от изображения с помощта на оптично разпознаване на символи (OCR) или други методи за обработка.
- **Зареждане и запазване на данни (Load)**: последната стъпка включва запазване на обработените данни в различни формати, като CSV, XLSX, XML, JSON, за по-нататъшна работа, като изборът на формат зависи от конкретните изисквания и предпочтения.

Този процес, известен като ETL (Extract, Transform, Load), играе ключова роля в автоматизираната обработка на данни и ще бъде разгледан по-подробно в главата "ETL и Pipeline: Extract, Transform, Load". По-нататък ще разгледаме примери за това как документи от различни формати се трансформират в структурирани данни.

### Пример за преобразуване на PDF -документ в таблица

Една от най-често срещаните задачи в строителните проекти е обработката на спецификации в PDF формат. За да демонстрираме прехода от неструктурирани данни към структуриран формат, нека разгледаме един практически пример: извлечане на таблица от PDF документ и преобразуването ѝ в CSV или Excel формат (фиг. 4.1-2).



Фиг. 4.1-2 За разлика от PDF, форматите CSV и XLSX са широко разпространени и лесно се интегрират в различни системи за управление на данни.

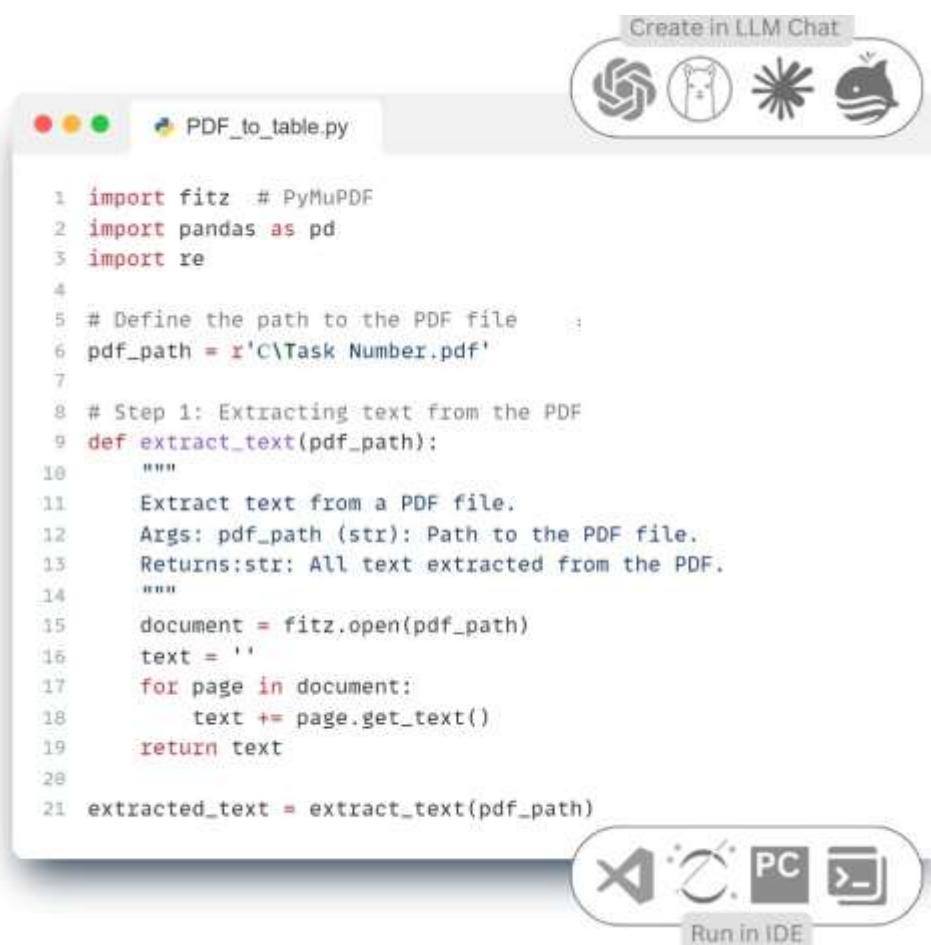
LLM езиковите модели, като ChatGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN значително опростяват начина, по който учените работят с данни, като намаляват необходимостта от дълбоко изучаване на езици за програмиране и позволяват много задачи да се решават с текстови заявки.

Ето защо, вместо да прекарваме време в търсене на решения в интернет (обикновено на уебсайта StackOverFlow или в тематични форуми и чатове) или да се свързваме със специалисти по обработка на данни, можем да използваме възможностите на съвременните онлайн или местни LLM. Достатъчно е да зададем запитване и моделът ще предостави готов код за преобразуване на PDF-документ в табличен формат.

- ❷ Изпратете следната текстова заявка до всеки LLM -модел (CHATGPT, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN или друг):

Моля, напишете код за извлечане на текст от PDF -файл, който съдържа таблица.  
Кодът трябва да приеме пътя до файла като аргумент и да върне извлечената таблица като DataFrame

- Отговорът на LLM -модела в повечето случаи ще бъде под формата на код на езика Python , тъй като този език е широко използван за обработка на данни, автоматизация и работа с различни файлови формати:



The screenshot shows a user interface for generating Python code. At the top right is a button labeled "Create in LLM Chat". Below it are four circular icons representing different AI models: a neural network, a cat, a star, and a whale. The main area displays a Python script named "PDF\_to\_table.py". The code uses PyMuPDF and Pandas libraries to extract text from a PDF file. A detailed docstring is provided for the "extract\_text" function. At the bottom right is a "Run in IDE" button with icons for PyCharm, VS Code, Jupyter Notebook, and Spyder.

```

1 import fitz # PyMuPDF
2 import pandas as pd
3 import re
4
5 # Define the path to the PDF file
6 pdf_path = r'C\Task Number.pdf'
7
8 # Step 1: Extracting text from the PDF
9 def extract_text(pdf_path):
10     """
11     Extract text from a PDF file.
12     Args: pdf_path (str): Path to the PDF file.
13     Returns: str: All text extracted from the PDF.
14     """
15     document = fitz.open(pdf_path)
16     text = ''
17     for page in document:
18         text += page.get_text()
19     return text
20
21 extracted_text = extract_text(pdf_path)

```

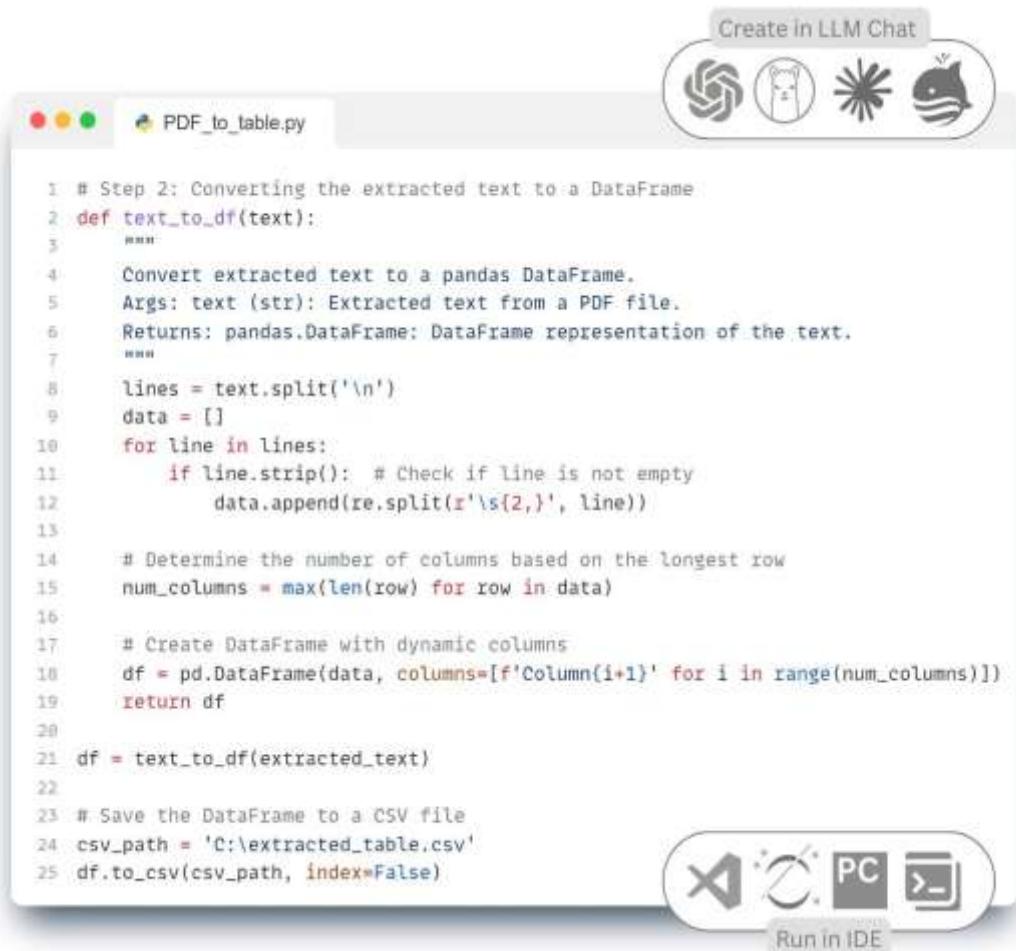
**Фиг. 4.1-3 Отговорът на LLM под формата на кода на Python и неговите библиотеки и пакети (Pandas, Fitz) извлича текст от PDF -файл.**

Този код (Фигура 4.1-3) може да бъде изпълнен оффлайн в една от популярните IDE, които споменахме по-горе: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с PyDev плъгин, Thonny, Wing IDE, IntelliJ IDEA с Python плъгин, JupyterLab или популярни онлайн инструменти: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

- В стъпката "Convert" (Конвертиране) използваме популярната библиотека Pandas (която разглеждахме подробно в главата "Python Pandas: незаменим инструмент за работа с данни"), за да прочетем извлечения текст в DataFrame и да запишем DataFrame във файл с CSV таблица или XLXS:

Нуждая се от код, който да преобразува получената таблица от PDF -файл в DataFrame. Също така да се добави код за запазване на DataFrame в CSV файл. ↵

💡 Отговор на LLM :



```

1 # Step 2: Converting the extracted text to a DataFrame
2 def text_to_df(text):
3     """
4         Convert extracted text to a pandas DataFrame.
5         Args: text (str): Extracted text from a PDF file.
6         Returns: pandas.DataFrame: DataFrame representation of the text.
7     """
8     lines = text.split('\n')
9     data = []
10    for line in lines:
11        if line.strip(): # Check if line is not empty
12            data.append(re.split(r'\s{2,}', line))
13
14    # Determine the number of columns based on the longest row
15    num_columns = max(len(row) for row in data)
16
17    # Create DataFrame with dynamic columns
18    df = pd.DataFrame(data, columns=[f'Column{i+1}' for i in range(num_columns)])
19    return df
20
21 df = text_to_df(extracted_text)
22
23 # Save the DataFrame to a CSV file
24 csv_path = 'C:\extracted_table.csv'
25 df.to_csv(csv_path, index=False)

```

Фиг. 4.1-4 Конвертиране на извлечената таблица от PDF в DataFrame и записване на таблицата в CSV -файл.

Ако по време на изпълнението на кода възникне грешка (фиг. 4.1-3, фиг. 4.1-4) - например поради липсващи библиотеки или грешен път на файла - текстът на грешката може просто да се копира заедно с изходния код и да се изпрати отново в LLM -модела. Моделът ще анализира съобщението за грешка, ще обясни какъв е проблемът и ще предложи поправки или допълнителни стъпки.

По този начин взаимодействието с AI LLM се превръща в завършен цикъл: за-явл→→тест→ обратна връзка→ корекция - без да са необходими задълбочени технически познания.

С помощта на обикновена текстова заявка в LLM chat и десетина реда на Python, които можем да стартираме локално във всяка IDE, преобразувахме PDF -документ в табличен CSV формат, който, за разлика от PDF документа, е лесно четим от машината и бързо се интегрира във всяка система за управление на данни.

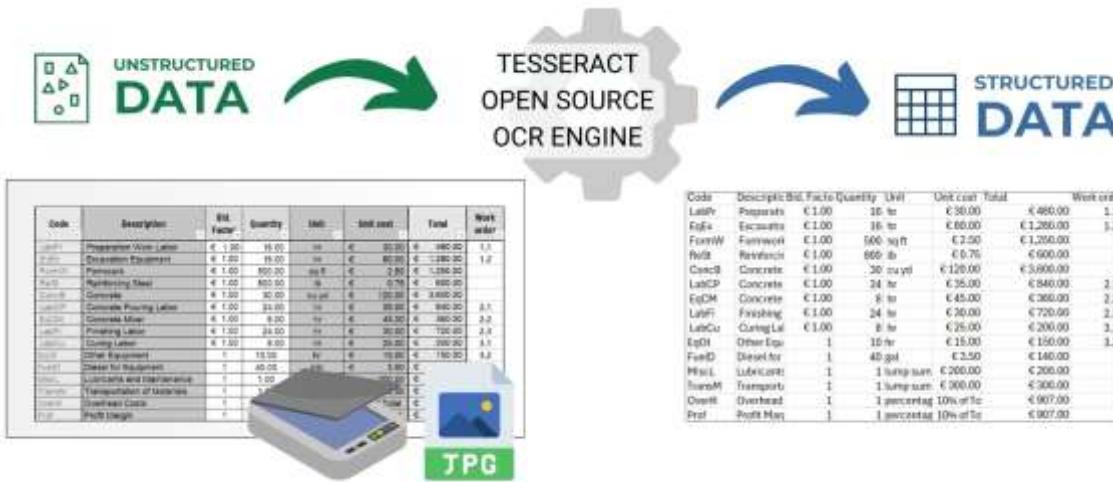
Можем да приложим този код (Фигура 4.1-3, Фигура 4.1-4), като го копираме от всеки чат на LLM, към десетки или хиляди нови PDF документи на сървъра, като по този начин автоматизираме процеса на преобразуване на поток от неструктурирани документи в структуриран формат на CSV таблица.

Но PDF документите невинаги съдържат текст, по-често това са сканирани документи, които трябва да бъдат обработени като изображения. Въпреки че изображенията по своята същност са неструктурирани, разработването и прилагането на библиотеки за разпознаване ни позволява да извличаме, обработваме и анализираме тяхното съдържание, което ни дава възможност да използваме пълноценно тези данни в нашите бизнес процеси.

## Конвертиране на JPEG, PNG изображение в структурирана форма

Изображенията са една от най-разпространените форми на неструктурирани данни. В строителството и в много други отрасли огромно количество информация се съхранява под формата на сканирани документи, схеми, снимки и чертежи. Такива данни съдържат ценна информация, но не могат да бъдат директно обработени, например като електронна таблица в Excel или база данни. Изображенията съдържат много сложна информация, тъй като тяхното съдържание, цветове, текстури са разнообразни и за извлечането на полезна информация е необходима специална обработка.

Трудността при използването на изображенията като източник на данни се състои в липсата на структура. Изображенията не предават значението по прям, лесно измерим начин, който компютърът може да разбере или обработи веднага, както това прави електронна таблица в Excel или таблица в база данни. За да се преобразуват неструктуриранныте данни от изображения в структурирана форма, трябва да се използват специални библиотеки, способни да интерпретират съдържащата се в тях визуална информация (фигура 4.1-5).



Фиг. 4.1-5 Конвертирането на сканирани документи и изображения в структурирани формати е възможно със специални инструменти за OCR.

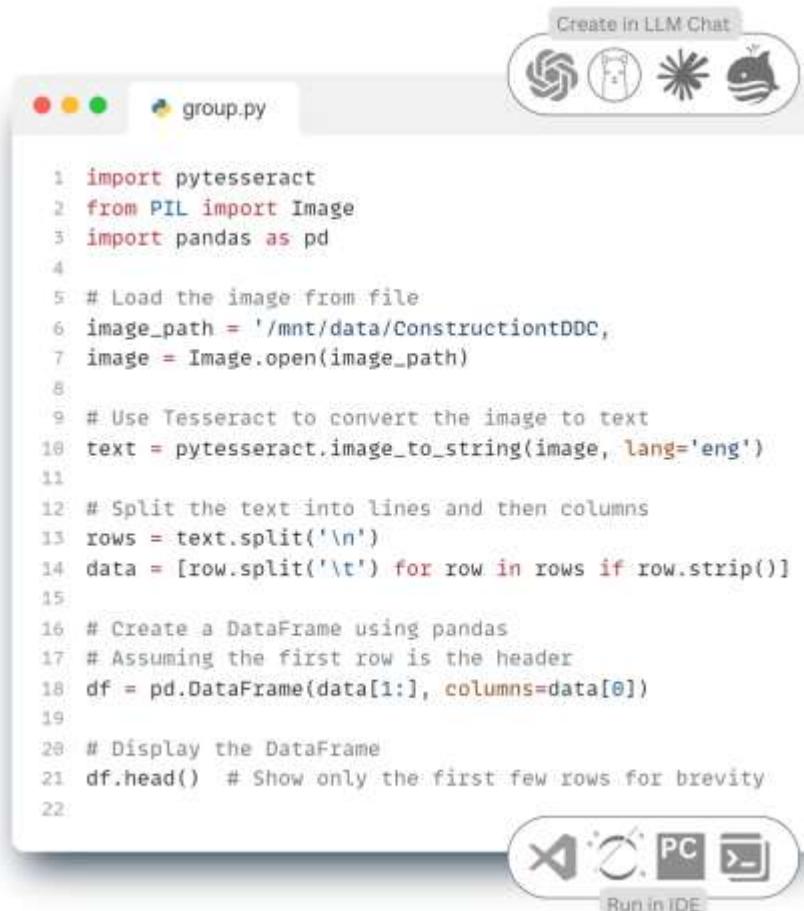
Технологията OCR (Optical Character Recognition - оптично разпознаване на символи) се използва за извличане на текст от изображения. Тя ви позволява да разпознавате букви и цифри в сканирани документи, снимки и PDF -файлове, като ги превръща в текст, който може да се редактира и чете от машини. Технологиите за OCR отдавна се използват в автоматизацията на документи, а днес тези лесно се интегрират във всякакви бизнес процеси и Python -приложения. Един от най-популярните инструменти за OCR е Tesseract, софтуер с отворен код, първоначално разработен от HP™ и сега поддържан от Google™. Той поддържа над 100 езика и има висока точност на разпознаване.

Нека помолим LLM chat да напише примерен код за извличане на данни от сканирана или фотографирана таблица по структуриран начин.

- Изпратете текстова заявка до чата LLM (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN или някой друг):

Напишете код за преобразуване на JPEG изображение, съдържащо таблица, в таблица DataFrame ↴

- В отговора на LLM в повечето случаи се предлага да се използва библиотеката Pytesseract за разпознаване на текст в изображения:



```

1 import pytesseract
2 from PIL import Image
3 import pandas as pd
4
5 # Load the image from file
6 image_path = '/mnt/data/ConstructionDDC,
7 image = Image.open(image_path)
8
9 # Use Tesseract to convert the image to text
10 text = pytesseract.image_to_string(image, lang='eng')
11
12 # Split the text into lines and then columns
13 rows = text.split('\n')
14 data = [row.split('\t') for row in rows if row.strip()]
15
16 # Create a DataFrame using pandas
17 # Assuming the first row is the header
18 df = pd.DataFrame(data[1:], columns=data[0])
19
20 # Display the DataFrame
21 df.head() # Show only the first few rows for brevity
22

```

**Фиг. 4.1-6 Конвертиране на текст, извлечен от таблица с изображения или снимки, в структурирано таблично представяне.**

В този пример - кодът (Фигура 4.1-6), получен от LLM, използва библиотеката pytesseract (Tesseract за Python) за преобразуване на изображение в текст с помощта на OCR (оптично разпознаване на символи) и библиотеката Pandas за преобразуване на този текст в структурирана форма, т.е. DataFrame.

Процесът на преобразуване обикновено включва предварителна обработка за подобряване на качеството на изображението, след което се прилагат различни алгоритми за откриване на модели, извлечане на характеристики или разпознаване на обекти. В резултат на това неструктурраната визуална информация се преобразува в структурирани данни.

Въпреки че PDF и изображенията са ключови източници на неструктуррана информация, истинският шампион по обем е текстът, генериран в имейли, чатове, срещи, месинджъри. Тези данни са

не просто многобройни - те са разпръснати, неформализирани и изключително зле структурирани.

## Конвертиране на текстови данни в структурирана форма

В допълнение към PDF документите с таблици (фиг. 4.1-2) и сканираните версии на таблични форми (фиг. 4.1-5), значителна част от информацията в проектната документация е представена в текстова форма. Тя може да бъде както последователни изречения в текстови документи, така и фрагментарни записи, разпръснати върху чертежи и схеми. В съвременните условия на обработка на данни една от най-често срещаните задачи е да се преобразува такъв текст в структуриран формат, подходящ за анализ, визуализация и вземане на решения.

Централно място в този процес заема таксономията - система за класификация, която позволява информацията да бъде организирана в категории и подкатегории въз основа на общи характеристики.

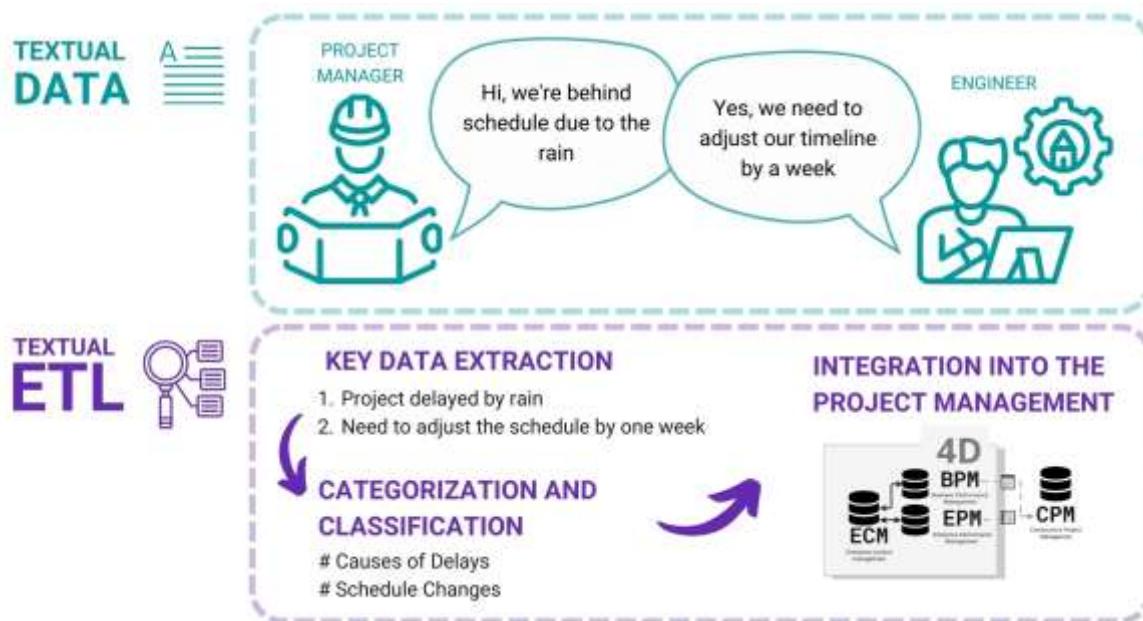
**Таксономията** е йерархична класификационна структура, която се използва за групирание и организиране на обекти. В контекста на обработката на текстове тя служи като основа за систематично категоризиране на елементите в семантични категории, като по този начин опростява анализа и подобрява качеството на обработката на данни.

Създаването на таксономия се съпровожда от етапите на извлечение на същности, категоризиране и контекстуализиране. За да моделираме процеса на извлечение на информация от текстови данни, трябва да следваме следните стъпки, които са подобни на тези, които вече приложихме за структуриране на данни от PDF документи:

- Извличане на информация): трябва да анализирате текстови данни, за да извлечете информация за закъснения и промени в графика на проекта.
- **Категоризиране и класифициране** (Transofrm): категоризиране на получената информация, например причини за закъснения и промени в графика.
- **Интеграция** (Load): накрая подгответе структурираните данни за интеграция във външни системи за управление на данни.

Разгледайте следната ситуация: имаме диалог между ръководител на проект и инженер, в който се обсъждат проблеми с графика. Нашата цел е да извлечем ключовите елементи (причини за забавяне, корекции в графика) и да ги представим по структуриран начин (фигура 4.1-7).

Нека да извършим извлечането въз основа на очакваните ключови думи, да създадем DataFrame, за да симулираме извлечането на данни, и след преобразуването - нова таблица DataFrame, която ще съдържа колони за дата, събитие (напр. причина за забавяне) и действие (напр. промяна в графика).



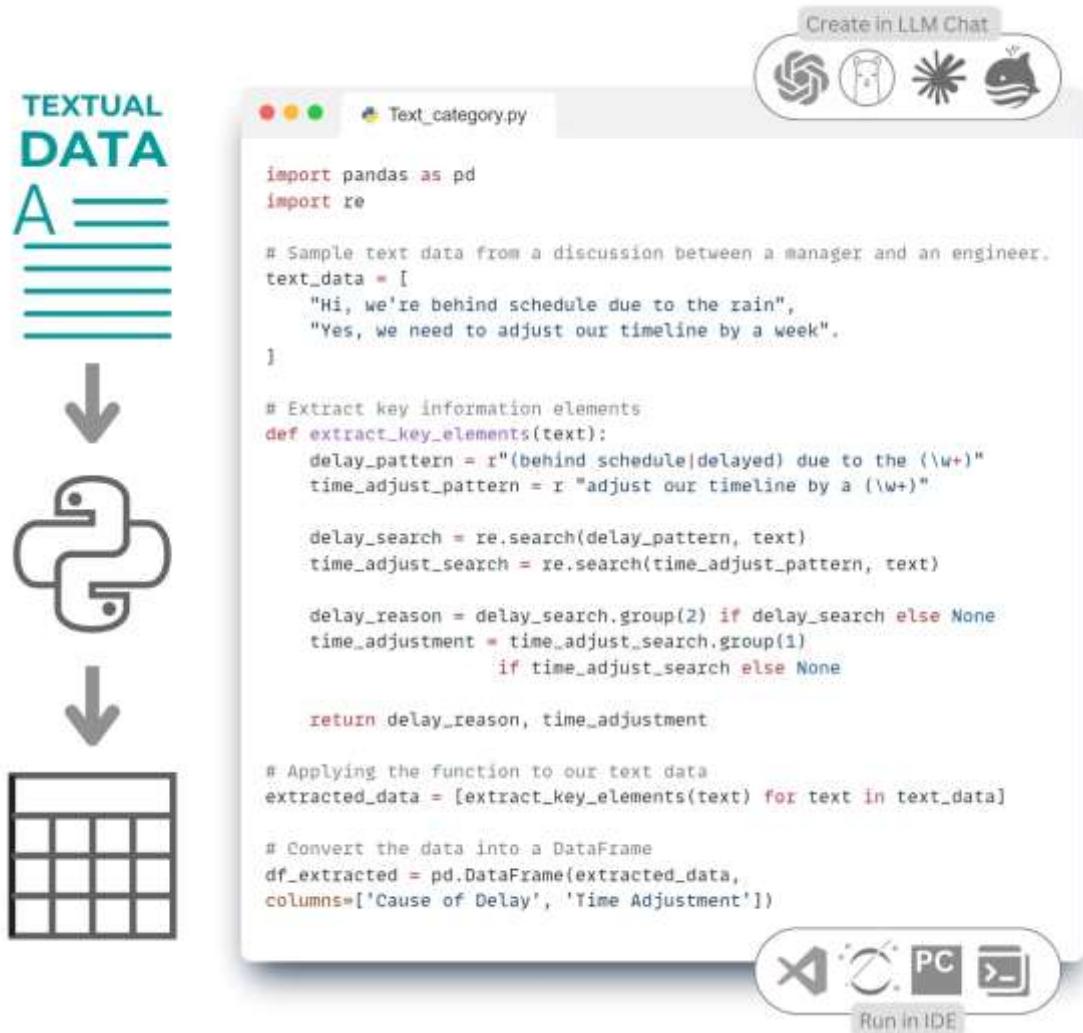
**Фигура 4.1-7** Подчертаване на ключова информация от текста за необходимостта от коригиране на сроковете и интегриране на промените в системата за управление на проекта.

Ето кода за решаване на задачата, като се използва текстова заявка в един от езиковите модели, както в предишните примери.

- ❷ Изпратете текстова заявка до всяка чат стая на LLM:

Имам разговор между мениджър: "Здравейте, изоставаме от графика заради дъждъ" и инженер: "Да, трябва да коригираме крайния срок с една седмица". Нуждая се от скрипт, който да анализира бъдещи подобни текстови диалози, да извлича от тях причините за закъсненията и необходимите корекции на крайния срок и след това да генерира DataFrame от тези данни. След това DataFrame трябва да се запише в CSV -файл. ↴

- 2 Отговорът от LLM обикновено включва Python -код, използващ регулярни изрази (re - Regex) и библиотеката Pandas (pd) :

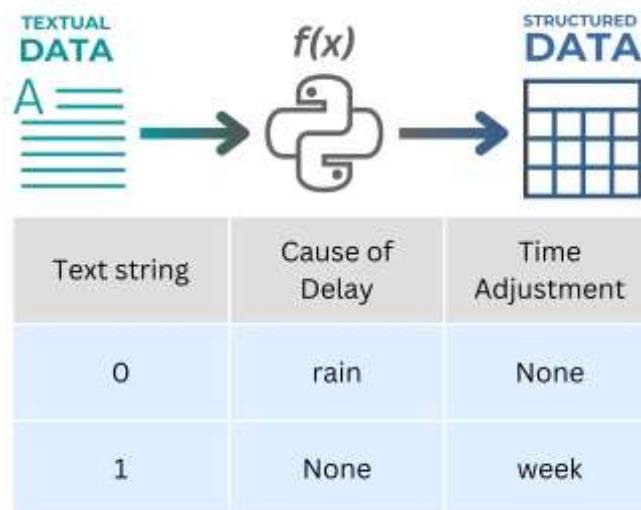


Фигура 4.1-8 Подчертаване на ключова информация от текста относно необходимостта от коригиране на крайните срокове в таблица.

В този пример (фигура 4.1-7) се анализират текстови данни, съдържащи кореспонденция между ръководител на проект и инженер, за да се идентифицира и извлече специфична информация, която може да повлияе на управлението на бъдещи проекти с подобни диалози. С помощта на регулярни изрази (ще обсъдим повече за регулярните изрази в главата "Структурирани изисквания и RegEx регулярни изрази"), чрез шаблони се идентифицират причините за закъсненията на проекта и необходимите корекции в графика. Функцията, написана в този пример, извлича или причината за закъснението, или корекцията във времето от низовете въз основа на шаблоните: избира думата след "поради" като причина за закъснението или думата след "от" като корекция във времето.

Ако в даден ред се споменава закъснение, дължащо се на метеорологични условия, като причина

се определя "дъжд"; ако в даден ред се споменава корекция на графика за определен период, този период се извлича като корекция на времето (фигура 4.1-9). Липсата на която и да е от тези думи в даден ред води до стойност "Няма" за съответната колона с атрибути.



**Фиг. 4.1-9** Обобщаващата таблица, получена като DataFrame, след изпълнение на кода, съдържа информация за наличието на закъснения и необходимите корекции на времето.

Структурирането и параметризирането на условията от текста (диалог, писмо, документ) позволява бързо да се елиминират закъсненията в строителството: например, липсата на работници може да повлияе на темпото на работа прилошо време, така че компаниите, които знаят параметрите на закъснението от диалозите (фиг. 4.1-9) между бригадира на строителния обект и ръководителя на проекта - предварително могат да подсияят екипа в случай на неблагоприятна прогноза.

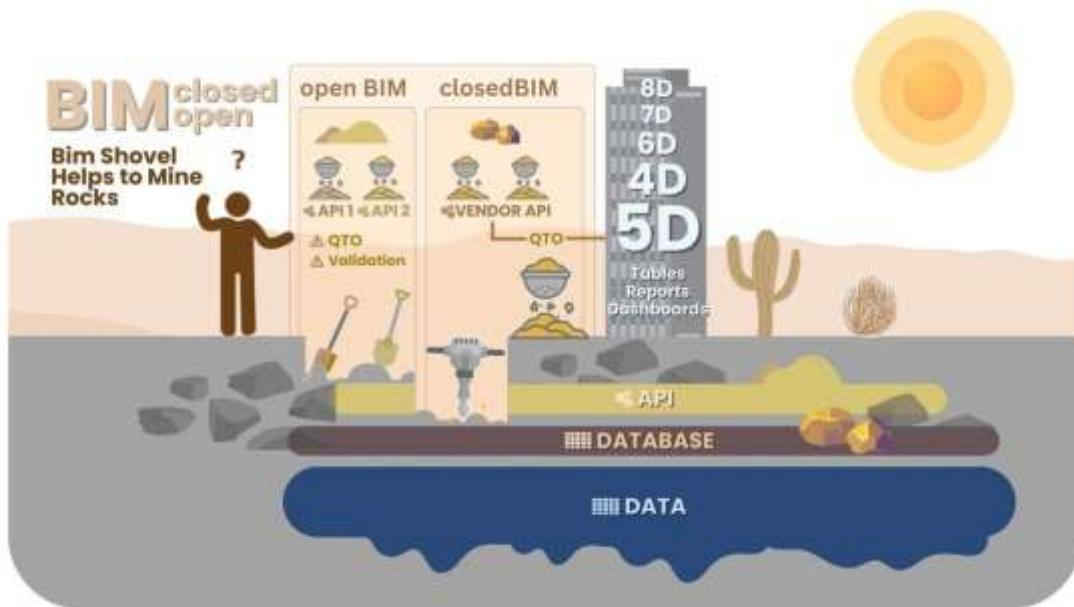
Конвертирането на документи и изображения в структуриран формат може да се осъществи с помощта на сравнително прости, отворени и бесплатни инструменти, базирани на категоризация.

Категоризацията на елементите също е ключова част от работата с данни за проекта, особено в контекста на използването на CAD софтуер (BIM).

## Конвертиране на CAD данни (BIM) в структуриран вид

Структурирането и категоризирането на CAD данни (BIM) е по-сложна задача, тъй като данните, съхранявани от CAD (BIM) базите данни, почти винаги са в затворени или сложни параметрични формати, като често съчетават едновременно елементи на геометрични данни (полуструктурирани) и елементи на метаинформация (полуструктурирани или структурирани данни).

Родните формати на данни в CAD (BIM) системите обикновено са защитени и недостъпни за директна употреба, освен ако не се използва специализиран софтуер или API - интерфейси на самия разработчик (фиг. 4.1-10). Тази изолация на данните формира затворени силози за съхранение, които ограничават свободния обмен на информация и възпрепятстват създаването на цялостни цифрови процеси в компанията.



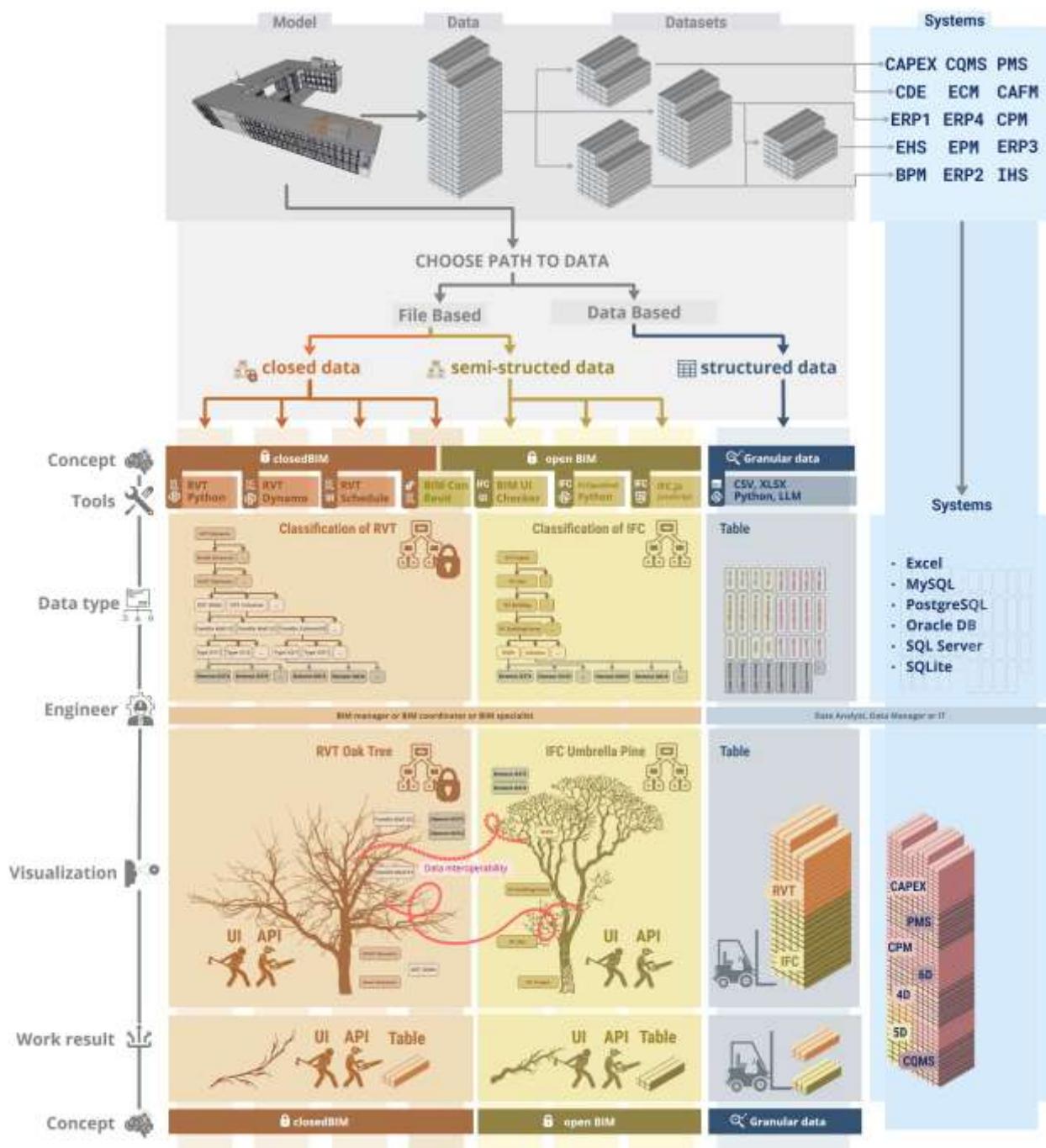
Фигура 4.1-10 CAD специалистите (BIM) могат да получат достъп до местни данни чрез API -връзки или инструменти на доставчика.

В специалните формати на CAD (BIM) информацията за характеристиките и атрибутите на елементите на проекта се събира в юерархична класификационна система, в която единиците със съответните свойства са разположени, подобно на плодовете на плодно дърво, в най-новите възли на клоните на класификацията на данните (фиг. 4.1-11).

Извличането на данни от такива юерархии е възможно по два начина: или ръчно, като се щракне върху всеки възел, сякаш се обработва дърво, като се отсекат избрани клони на категории и типове с брадва. Алтернативно, използването на интерфейси за програмиране на приложения (API) осигурява по-ефективен, автоматизиран подход за извлечане и групиране на данни, като в крайна сметка ги трансформира в структурирана таблица за използване в други системи.

За извличане на структурирани таблици с данни от CAD (BIM) проекти могат да се използват различни инструменти, като Dynamo, pyRvt, Pandamo (Pandas + Dynamo), ACC или решения с отворен код, като IfcOpSh или IFCjs за IFC формат.

Съвременните инструменти за експортиране и конвертиране на данни дават възможност, с цел опростяване на обработката и подготовката на данните, да се раздели съдържанието на CAD моделите на два основни компонента: геометрична информация и данни за атрибути (фиг. 4.1-13) - метаинформация, описваща свойствата на елементите на дизайна (фиг. 3.1-16). Тези два слоя данни остават свързани чрез уникални идентификатори, благодарение на които е възможно точното сътресяне на всеки елемент с описание на геометрията (чрез параметри или полигони) към неговите атрибути: име, материал, етап на завършеност, цена и т.н. Този подход гарантира целостта на модела и позволява гъвкаво използване на данните както за визуализация (данни за геометричния модел), така и за аналитични или управлениски задачи (структурирани или свободно структурирани), като се работи с двата вида данни поотделно или паралелно.



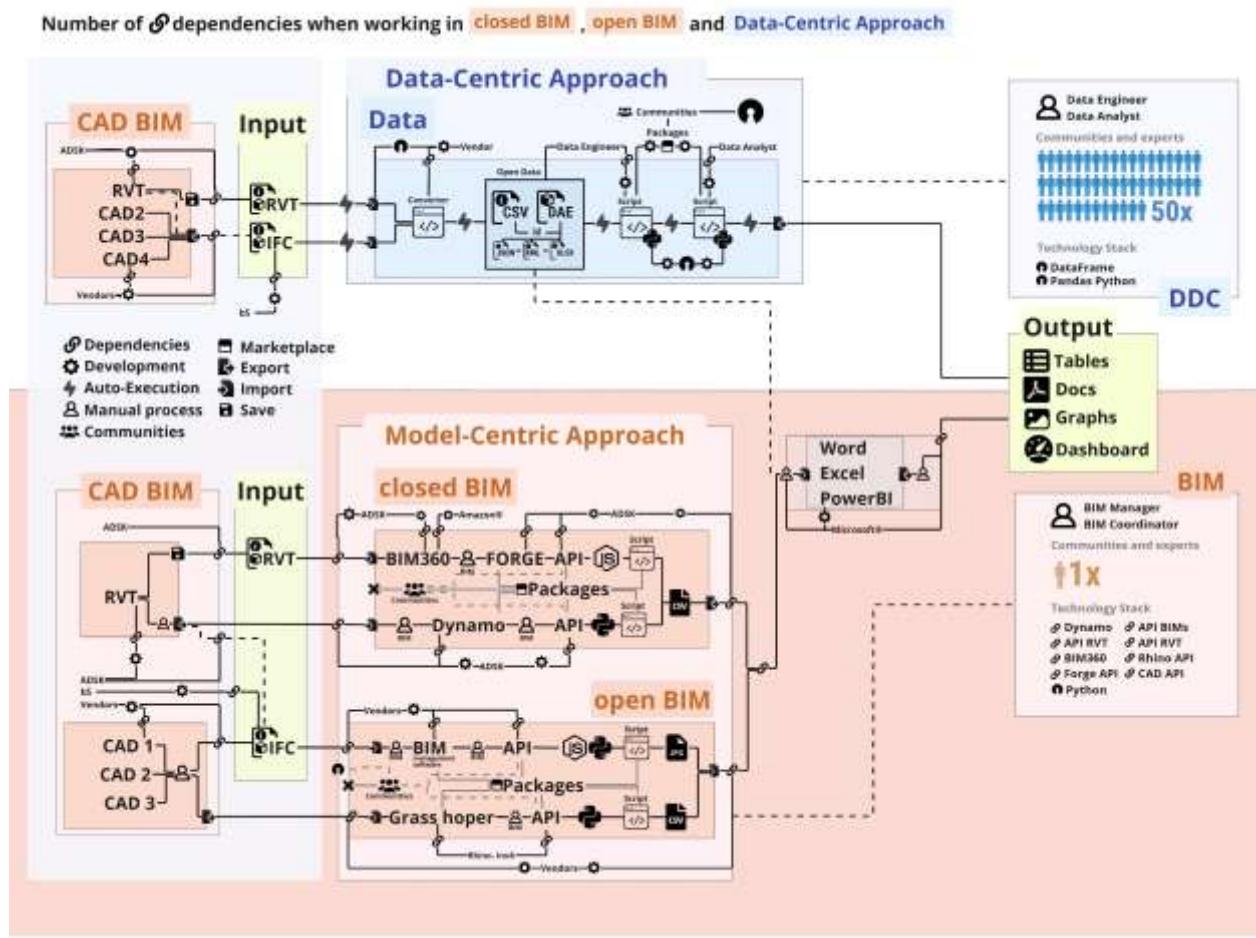
Фиг. 4.1-11 Изгледът на информацията от CAD базите данни (BIM) се представя на потребителя под формата на класификационни дървета.

С развитието на технологията за обратен инженеринг и появата на SDK (Software Development Kit) за конвертиране на CAD данни - наличието и конвертирането на данни от затворени CAD софтуерни формати (BIM) стана много по-лесно. Вече е възможно законно и безопасно да се конвертират данни от затворени формати в универсални формати, подходящи за анализ и използване в други

системи. Историята на първите инструменти за обратен инженеринг ("Open DWG") и борбата за господство над форматите на производителите на CAD беше разгледана в главата "Структурирани данни: основа на цифровата трансформация".

Инструментите за обратен инженеринг позволяват легитимно извлечение на данни от затворени патентовани формати, като разбиват информацията от смесения формат на CAD (BIM) на типове данни и формати, изисквани от потребителя, което улеснява обработката и анализа.

С помощта на обратен инженеринг и директен достъп до информация от CAD базите данни прави информацията достъпна, като позволява използването на отворени данни и отворени инструменти, както и анализ на данните с помощта на стандартни инструменти, изготвяне на отчети, визуализации и интеграция с други цифрови системи (фиг. 4.1-12).



Фиг. 4.1-12 Директният достъп до CAD данни свежда до минимум зависимостта от софтуерни платформи и преминава към подход, ориентиран към данните.

От 1996 г. за формата DWG, от 2008 г. за формата DGN и от 2018 г. за RVT е възможно да се конвертират първоначално затворени CAD формати данни във всякакви други формати, включително структурирани формати, удобно и ефективно с помощта на инструменти за обратно инженерство

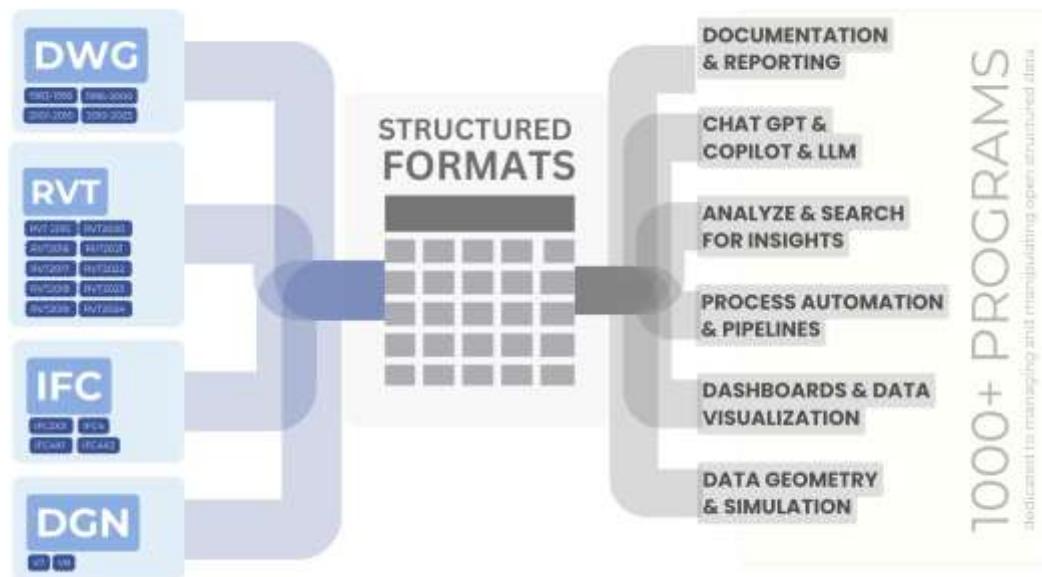
(фиг. 4.1-13). Днес почти всички големи CAD (BIM) и големи инженерни компании в света използват SDK - инструменти за обратно инженерство, за да извличат данни от затворени CAD (BIM) формати на доставчиците [92].



**Фиг. 4.1-13** Използването на инструменти за обратен инженеринг позволява базите данни на CAD (BIM) програмите да бъдат преобразувани във всеки удобен модел на данни.

Конвертирането на данни от затворени, патентовани формати в отворени формати и разделянето на смесени CAD (BIM) формати на геометрични данни и данни с метаинформационни атрибути опростява процеса на работа с тях, като ги прави достъпни за анализ, манипулиране и интегриране с други системи (фиг. 4.1-14).

При днешната работа с CAD данни (BIM) сме достигнали до момент, в който не е необходимо да искаме разрешение от доставчиците на CAD (BIM) за достъп до информация от CAD формати.



**Фиг. 4.1-14** Съвременните SDK инструменти позволяват легално конвертиране на данни от патентовани формати на CAD бази данни (BIM).

Настоящите тенденции в обработката на данни за CAD проектиране продължават да се определят от ключови участници на пазара - доставчици на CAD - които работят за укрепване на позициите си в света на данните и създават нови формати и концепции.

## Доставчиците на CAD решения се насочват към структурирани данни

От 2024 г. нататък в сектора на проектирането и строителството предстои значителна технологична промяна в използването и обработката на данни. Вместо свободен достъп до данни за проектиране, доставчиците на CAD -системи се фокусират върху популяризирането на следващите нови концепции. Подходи като BIM (създаден през 2002 г.) и отворен BIM (създаден през 2012 г.) постепенно отстъпват място на съвременни технологични решения, които доставчиците на CAD системи започват да популяризират [93]:

- Преминаване към използване на "гранулирани" данни, което позволява ефективно управление на информацията и преминаване към анализ на данните.
- Възникване на формата USD и прилагане на подхода Entity-component-system (ECS) за гъвкава организация на данните
- Активно използване на изкуствен интелект при обработка на данни, автоматизация на процеси и анализ на данни
- Разработване на оперативна съвместимост - подобряване на взаимодействието между различни програми, системи и бази данни

Всеки от тези аспекти ще бъде разгледан по-подробно в шестата част на книгата "CAD и BIM: маркетинг, реалност и бъдеще на данните за проектиране в строителството". В рамките на тази глава само накратко ще очертаем общия вектор на промяната: основните доставчици на CAD сега се стремят да преосмислят начина, по който се структурира информацията за проектирането. Една от ключовите промени е изоставянето на класическия модел за съхранение на данни, базиран на файлове, в полза на гранулирана, аналитично ориентирана архитектура на данните, която осигурява непрекъснат достъп до отделните компоненти на модела [93].

Същността на случващото се е, че индустрията постепенно се отказва от тромавите, специализирани и параметрични формати, които изискват геометрични ядра, в полза на по-универсални, машинно четими и гъвкави решения.

Един такъв двигател на промяната е форматът USD (Universal Scene Description), първоначално разработен в индустрията за компютърна графика, но вече признат в инженерните приложения благодарение на разработването на платформата NVIDIA Omniverse (и Isaac Sim) за симулации и визуализации [93]. За разлика от параметричния IFC, USD предлага по-опростена структура и позволява описание на геометрията и свойствата на обектите във формат JSON (фигура 4.1-15), което улеснява обработката на информацията и ускорява интегрирането ѝ в цифровите процеси. Новият формат позволява съхраняване на геометрия (в допълнение към BREP -NURBS - повече подробности в част 6 на книгата) под формата на MESH полигона, а свойствата на обектите - в JSON, което го прави по-удобен за автоматизирани процеси и работа в облачни екосистеми [94].

Някои производители на CAD и ERP вече използват подобни формати (например NWD, SVF, CP2,

CPIXML), но повечето от тях остават затворени и недостъпни за външна употреба, което ограничава възможностите за интегриране на данни и повторно използване. В този контекст USD може да играе същата роля, каквато е играл навремето DXF - отворена алтернатива на патентовани формати като DWG.

General Information				Comparison / Notes
Year of format creation	1991	2016		IFC focuses on construction data, USD on 3D graphics
Creator-developer	TU Munich	Pixar		IFC was founded in Germany, USD in America
Prototypes and predecessors	IGES, STEP	PTEX, DAE, GLTF		IFC evolved from IGES/STEP, USD from PTEX/DAE/GLTF
Initiator in Construction	ADSK	ADSK		ADSK initiated the adoption of both formats in construction
Organizer of the Alliance	ADSK	ADSK		ADSK organized both alliances
Name of the Alliance	bS (IAI)	AOUSD		Different alliances for each format
Year of Alliance Formation	1994	2023		The IFC alliance was formed in 1994, AOUSD for USD in 2023
Promoting in the construction	ADSK and Co	ADSK and Co		ADSK and Co actively promotes both formats in bS (IAI) since the introduction

Purpose and Usage				Comparison / Notes
Purpose	Semantic description and interoperability	Data simplification, visualization unification		IFC for semantics and exchange; USD for simplification and visualization
Goals and Objectives	Interoperability and semantics	Unification for visualization and data processing		IFC focuses on semantics; USD on visualization
Use in Other Industries	Predominantly in construction	In film, games, VR/AR, and now in construction		USD is versatile and used in various fields
Supported Data Types	Geometry, object attributes, metadata	Geometry, shaders, animation, light, and camera		USD supports a wider range of data types suitable for complex visualizations; IFC focuses on construction-specific data

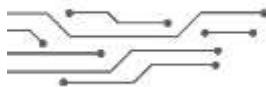
Фиг. 4.1-15 USD формат като опит на производителите на CAD да отговорят на търсенето на оперативна съвместимост и независимост на данните за проектиране от геометричните ядра.

Преминаването на основните разработчици към отворени и опростени USD, GLTF, OBJ, XML (затворени NWD, CP2, SVF, SVF2, CPIXML) и други подобни формати (фиг. 3.1-17) отразява глобалната тенденция и търсенето на индустрията за опростяване на данните и увеличаване на достъпността. През следващите години можем да очакваме постепенно отдалечаване от сложните параметрични стандарти и формати със зависимост от геометрични ядра в полза на по-леки и структурирани решения. Този преход ще ускори цифровизацията на строителната индустрия, ще улесни автоматизацията на процесите и ще опрости обмена на данни.

Въпреки стратегическите планове на доставчиците на CAD за популяризиране на нови отворени формати, специалистите от строителния бранш могат да получат пълен достъп до данни от затворени CAD системи, без да се нуждаят от CAD (BIM) инструменти, като използват инструменти за обратно инженерство.

Всички тези тенденции неизбежно водят до преминаване от обемисти, монолитни 3D модели към универсални, структурирани данни и към използване на формати, които отдавна са се доказали в други индустрии. След като проектните екипи започнат да възприемат CAD моделите не само като визуални обекти или набор от файлове, а като бази данни, съдържащи знания и информация, подходът към проектирането и управлението се променя драстично.

След като екипите се научат как да извлечат структурирани данни от документи, текстове, чертежи и CAD модели и имат достъп до бази данни, следващата ключова стъпка е моделирането на данни и осигуряването на качеството. Именно тази стъпка до голяма степен определя скоростта на обработка и преобразуване на информацията, която в крайна сметка ще се използва за вземане на решения в конкретни приложни задачи.



## ГЛАВА 4.2.

# КЛАСИФИКАЦИЯ И ИНТЕГРАЦИЯ: ОБЩ ЕЗИК ЗА СТРОИТЕЛНИТЕ ДАННИ

### **Бързината на вземане на решения зависи от качеството на данните**

Съвременната архитектура на данните за проектиране претърпява фундаментални промени. Индустрията се отдалечава от обемисти, изолирани модели и затворени формати към по-гъвкави, машинно четими структури, фокусирани върху анализа, интеграцията и автоматизацията на процесите. Преходът към нови формати обаче сам по себе си не гарантира ефективност - качеството на самите данни неизбежно е в центъра на вниманието.

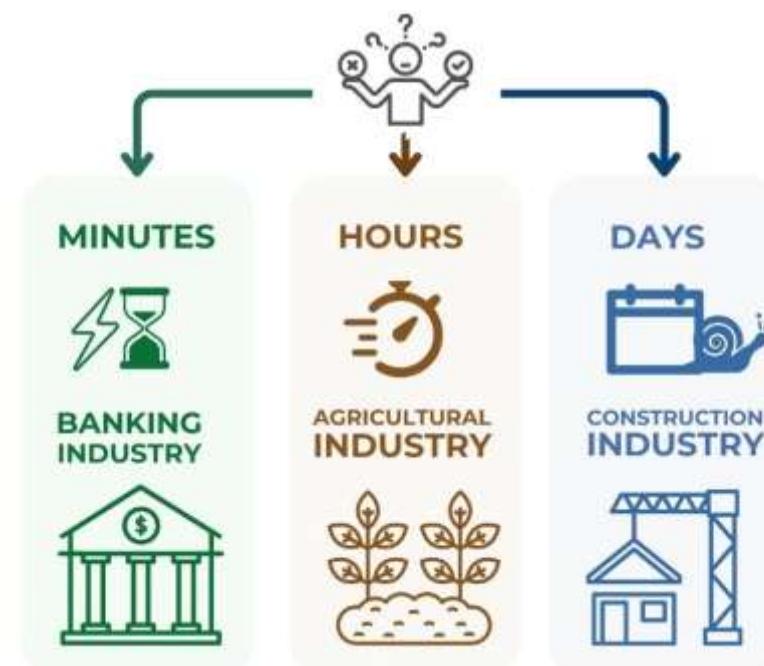
На страниците на тази книга говорим много за формати, системи и процеси. Но всички тези усилия са безсмислени без един ключов елемент: данни, на които може да се има доверие. Качеството на данните е крайъгълният камък на цифровизацията, към който ще се връщаме в следващите части.

Съвременните строителни компании - особено големите - използват десетки, а понякога и хиляди различни системи и бази данни (фиг. 4.2-1). Тези системи трябва не само да се попълват редовно с нова информация, но и да взаимодействат ефективно помежду си. Всички нови данни, генериирани в резултат на обработката на постъпващата информация, се интегрират в тези среди и служат за решаване на конкретни бизнес задачи.

И докато в миналото решенията по конкретни бизнес задачи са се вземали от топ мениджърите - т.нр. HiPPO (фиг. 2.1-9) - въз основа на опита и интуицията, днес, с рязкото увеличаване на обема на информацията, този подход става противоречив. Автоматизираният анализ, който работи с данни в реално време, го замества.

"Традиционно-ръчните" обсъждания на бизнес процесите на изпълнително ниво ще се изместят към оперативните анализи, които изискват бързи отговори на бизнес запитвания.

Епохата, в която счетоводители, майстори и оценители генерираха ръчно отчети, обобщаващи таблици и витрини с данни за проекта в продължение на дни и седмици, е в миналото. Днес бързината и навременността на вземането на решения се превръщат в ключов фактор за конкурентно предимство.



Фиг. 4.2-1 В строителния бранш изчисленията и вземането на решения отнемат дни, за разлика от други браншове, където това става за часове или минути.

Основната разлика между строителната индустрия и по-напредналите в цифрово отношение отрасли (фигура 4.2-1) е ниското ниво на качество и стандартизация на данните. Остарелите подходи към генерирането, предаването и обработката на информация забавят процесите и създават хаос. Липсата на единни стандарти за качество на данните възпрепятства внедряването на автоматизация от край до край

Едно от основните предизвикателства остава лошото качество на входящите данни, както и липсата на формализирани процеси за тяхната подготовка и валидиране. Без надеждни и последователни данни не може да има ефективна интеграция между системите. Това води до закъснения, грешки и увеличени разходи на всеки етап от жизнения цикъл на проекта.

В следващите раздели на книгата ще разгледаме подробно как можете да подобрите качеството на данните, да стандартизирате процесите и да съкратите пътя от информация до качествени, валидирани и последователни данни.

## Стандартизация и интеграция на данни

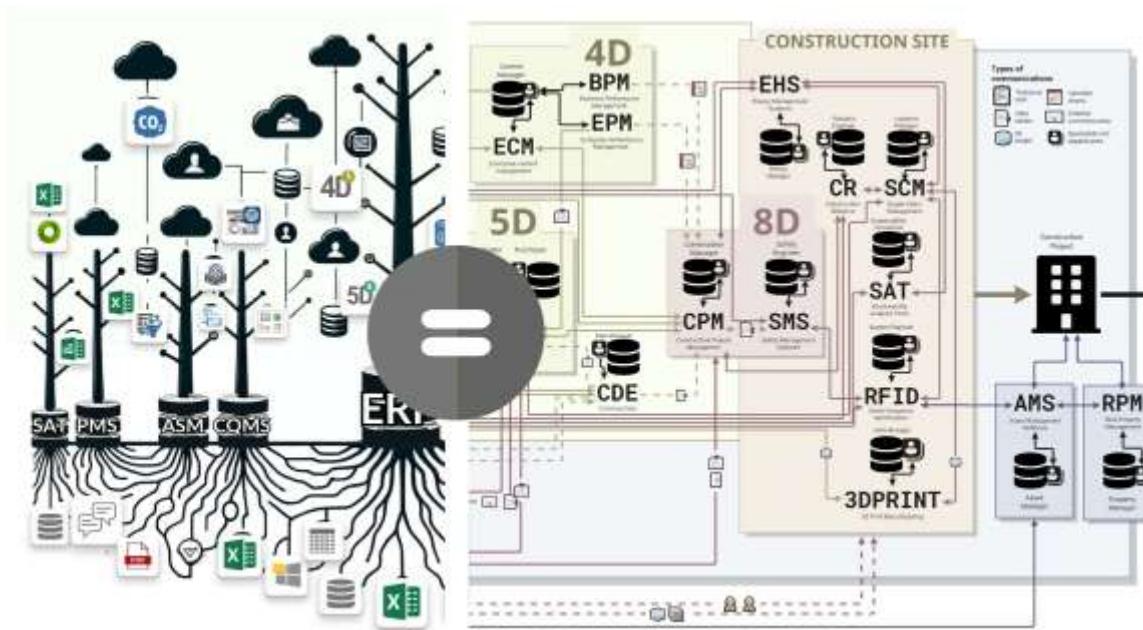
Ефективното управление на данните изисква ясна стратегия за стандартизация. Само с ясни изисквания за структурата и качеството на данните може да се автоматизира валидирането на данните, да се намалят ръчните операции и да се ускори вземането на информирани решения на всички етапи на проекта.

В ежедневната си практика строителната компания трябва да обработва стотици файлове всеки

ден: имейли, PDF -документи, CAD проектни файлове, данни от IoT сензори, които трябва да бъдат интегрирани в бизнес процесите на компанията.

Гората на екосистемата от бази данни и инструменти на компанията (фигура 4.2-2) трябва да се научи да извлича хранителни вещества от постъпващите многоформатни данни, за да постигне желаните от компанията резултати.

За да се справите ефективно с потока от данни, не е задължително да наемете армия от мениджъри, а първо трябва да разработите строги изисквания и стандарти за данните и да използвате подходящи инструменти за автоматичното им валидиране, хармонизиране и обработка.



**Фиг. 4.2-2** Осигуряването на здрава фирмена екосистема изиска качествено и своевременно ресурсно обезпечаване на нейните системи.

За да се автоматизира процесът на валидиране и хармонизиране на данните (за последваща автоматична интеграция), трябва да се започне с описание на минималните необходими изисквания за данните за всяка конкретна система. Тези изисквания определят:

- Какво точно трябва да получите?
- В каква форма (структура, формат)?
- Кои атрибути са задължителни?
- Какви допустими отклонения в точността и пълнотата са допустими?

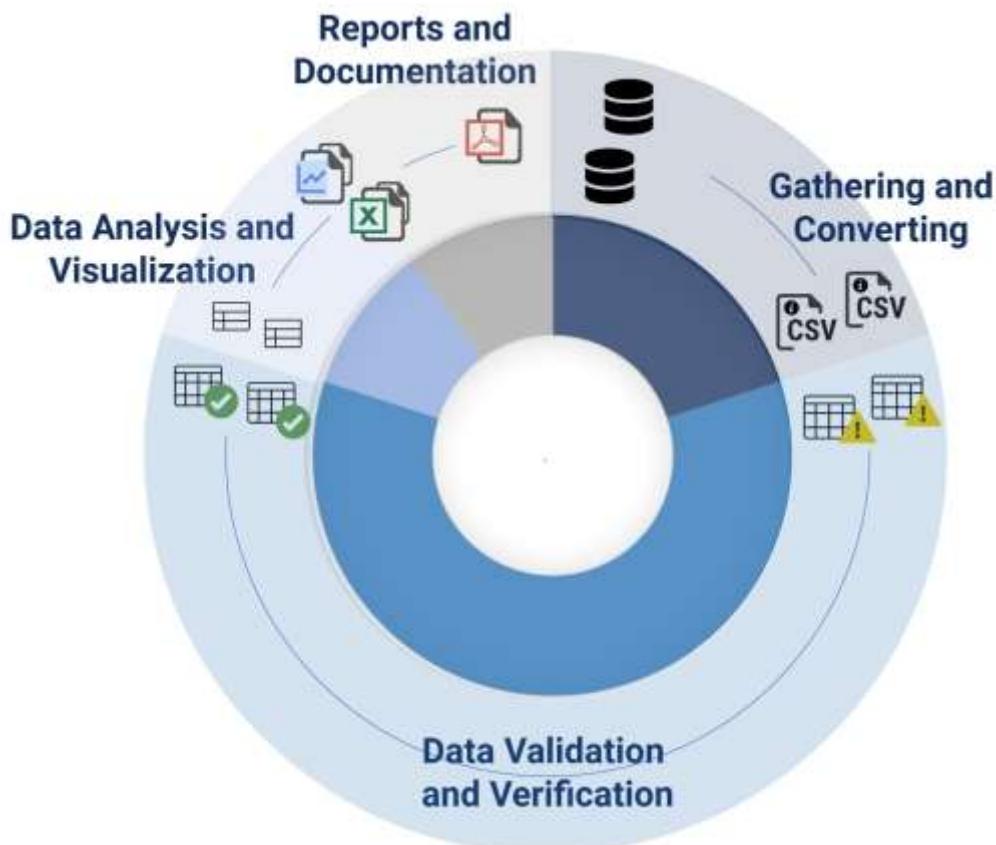
Изискванията към данните описват критериите за качеството, структурата и пълнотата на получаваната и обработваната информация. Например за текстовете в PDF -документи е важно те да бъдат точно форматирани в съответствие с индустриските стандарти (фиг. 7.2-14 - фиг. 7.2-16).

Обектите в CAD -моделите трябва да имат правилни атрибути (размери, кодове, връзки с класификатори) (фиг. 7.3-9, фиг. 7.3-10). А за сканирането на договори са важни ясните дати и възможността за автоматично извлечане на сумата и ключовите термини (фиг. 4.1-7 - фиг. 4.1-10).

Формулирането на изисквания за данни и автоматичната проверка на тяхното съответствие е една от най-времеемските, но важни стъпки. Това е най-времеемската стъпка в бизнес процесите.

Както беше споменато в част 3 на тази книга, между 50 и 90% от времето на специалистите по бизнес анализи (BI) се изразходва за подготовка на данни, а не за анализ (фигура 3.2-5). Този процес включва събиране на данни, проверка, валидиране, хармонизиране и структуриране.

Според проучване от 2016 г. [95] учените, занимаващи се с данни, в най-различни области с широк спектър на действие заявяват, че прекарват по-голямата част от работното си време (около 80%) в това, което най-малко обичат да правят (фигура 4.2-3): събиране на съществуващи набори от данни и тяхното организиране (обединяване, структуриране). По този начин по-малко от 20% от времето им остава за творчески задачи, като например намиране на модели и закономерности, които ще доведат до нови прозрения и открития.



Фиг. 4.2-3 Проверката на и осигуряването на качеството на данните е най-скъпата, времеемска и сложна стъпка при подготовката на данните за интегриране в други системи.

Успешното управление на данни в строителна компания изискава цялостен подход, който включва параметризиране на задачите, формулиране на изисквания за качество на данните и използване на подходящи инструменти за автоматичното им валидиране.

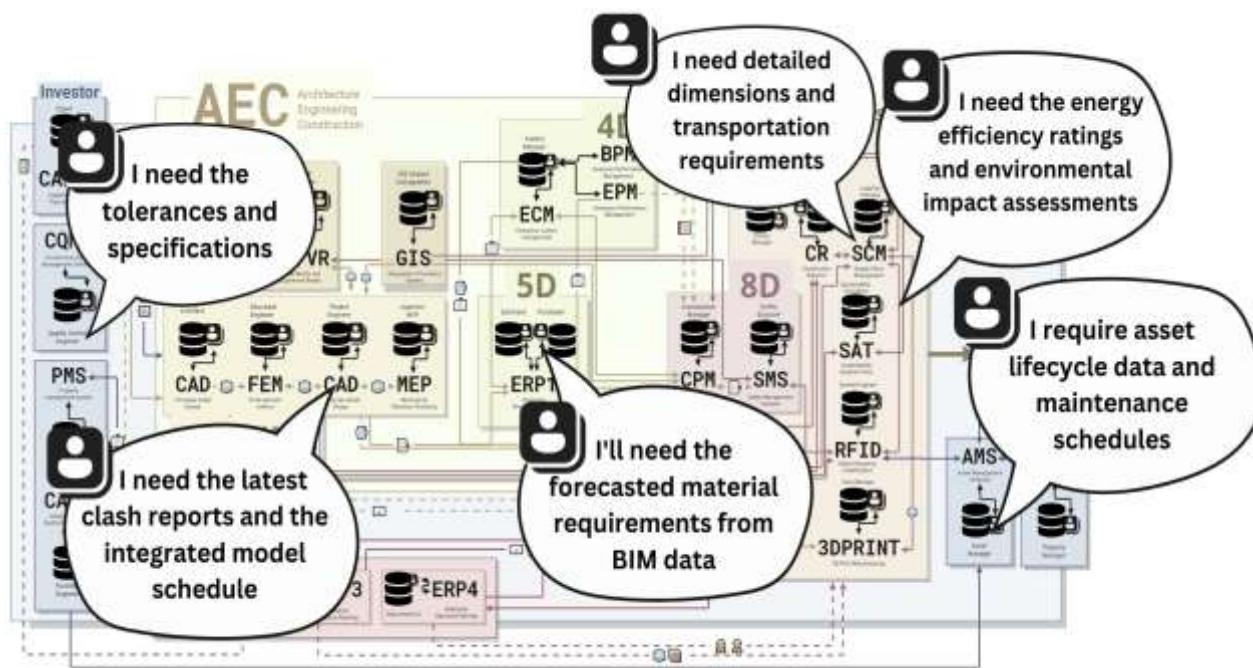
## Цифровата оперативна съвместимост започва с изискванията

С нарастването на броя на цифровите системи в компаниите нараства и необходимостта от съгласуваност на данните между тях. Мениджърите, отговарящи за различни ИТ системи, често се оказват неспособни да се справят с нарастващия обем информация и разнообразието от формати. При тези обстоятелства те са принудени да се обръщат към специалисти с молба да създадат данни във вид, подходящ за използване в други приложения и платформи.

Това от своя страна изиска от инженерите и служителите за генериране на данни да се адаптират към множество изисквания, често без прозрачност и ясно разбиране за това къде и как ще се използват данните в бъдеще. Липсата на стандартизиирани подходи за обработка на информацията води до неефективност и увеличаване на разходите по време на етапа на проверка, който често е ръчен поради сложността и нестандартизирания характер на данните.

Въпросът за стандартизацията на данните не е само въпрос на удобство или автоматизация. Той е пряка финансова загуба. Според доклад на IBM от 2016 г. годишните загуби от лошото качество на данните в САЩ са 3,1 трилиона долара [96]. Освен това проучвания на MIT и други аналитични консултантски фирми показват, че цената на лошото качество на данните може да достигне 15-25% от приходите на компанията [97].

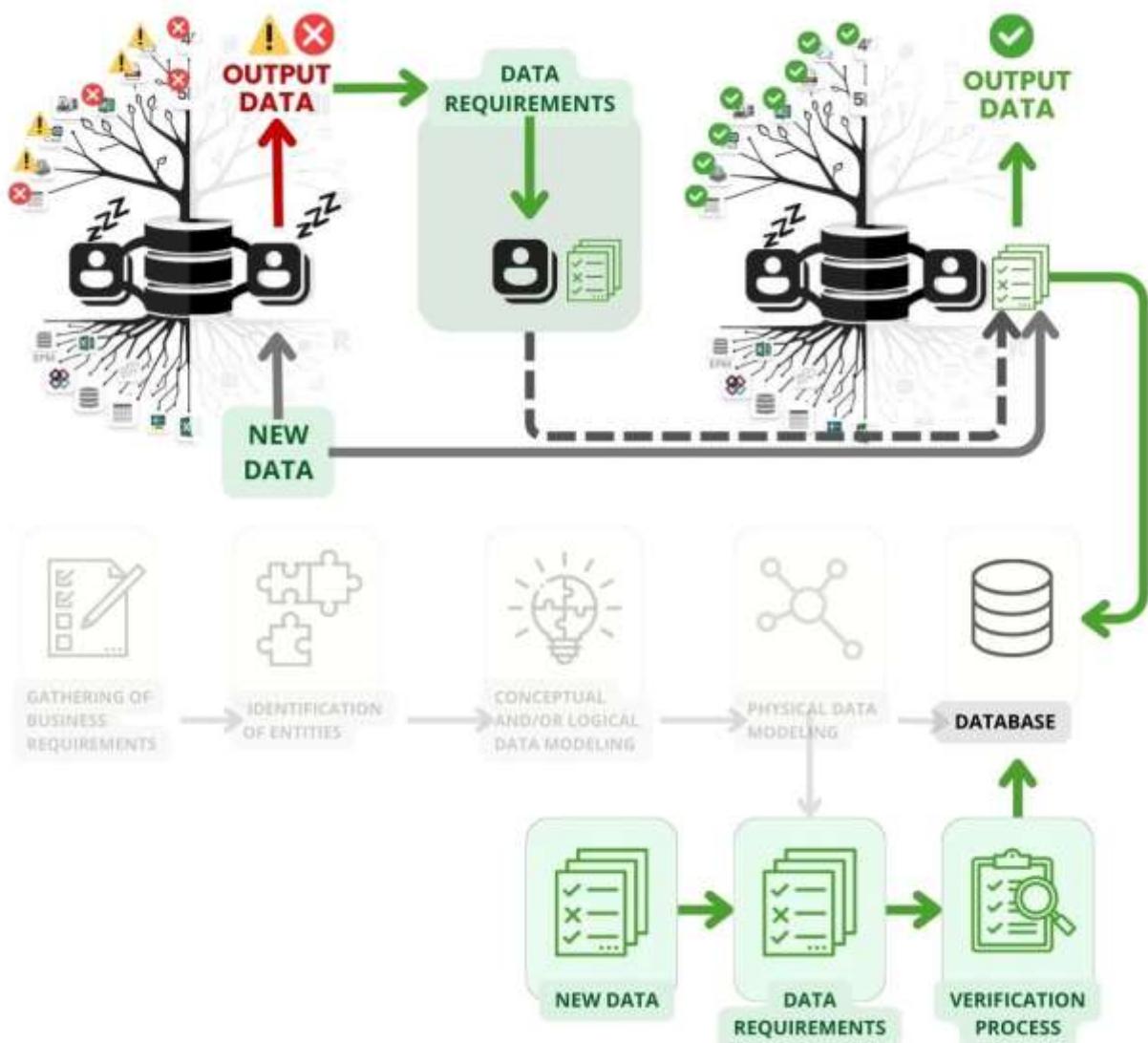
При тези условия става изключително важно да има ясно определени изисквания за данните и описания на това какви параметри, в какъв формат и с какво ниво на детайлност трябва да бъдат включени в създадените обекти. Без формализирането на тези изисквания е невъзможно да се гарантира качеството на данните и съвместимостта между системите и фазите на проекта (фиг. 4.2-4).



Фиг. 4.2-4 Бизнесът се основава на взаимодействието на различни роли, всяка от които изиска определени параметри и стойности, които са от решаващо значение за изпълнението на бизнес целите.

За да формулирате правилно изискванията към данните, трябва да разберете бизнес процесите на ниво данни. Строителните проекти се различават по вид, мащаб и брой участници и всяка система - независимо дали става въпрос за моделиране (CAD (BIM)), планиране (ERP 4D), калкулиране на разходите (ERP 5D) или логистика (SCM) - изиска свои собствени уникални параметри за входни данни (входни единици-елементи).

В зависимост от тези нужди бизнес мениджърите трябва или да проектират нови структури от данни, за да отговорят на изискванията, или да адаптират съществуващите таблици и бази данни. Качеството на създадените данни ще зависи пряко от това колко точно и правилно са формулирани изискванията (фиг. 4.2-5).



Фигура 4.2-5 Качеството на данните зависи от качеството на изискванията, които се създават за конкретни случаи на използване на данните.

Тъй като всяка система има свои специфични изисквания към данните, първата стъпка при формулирането на общите изисквания е да се категоризират всички елементи, участващи в бизнес процесите. Това означава необходимост от разделяне на обектите на класове и групи от класове, съответстващи на конкретни системи или задачи на приложението. За всяка такава група се разработват отделни изисквания за структурата, атрибутите и качеството на данните.

На практика обаче прилагането на този подход се сблъсква с голямо предизвикателство: липсата на общ език за групиране на данни. Различните класификации, дублиращите се идентификатори и несъвместимите формати водят до това, че всяка компания, всеки софтуер и дори всеки проект формират свои собствени, изолирани модели и класове данни. Резултатът е цифрова "Вавилонска кула", в която прехвърлянето на информация между системите изисква многократни преобразувания към правилните модели и класове данни, често извършвани ръчно. Тази бариера може да бъде

преодоляна само чрез преминаване към универсални класификатори и стандартизириани набори от изисквания.

## Общ език за конструиране: ролята на класификаторите в цифровата трансформация

В контекста на цифровизацията и автоматизацията на процесите на инспекция и обработка системите за класификация играят специална роля - един вид "цифрови речници", които осигуряват единство при описанието и параметризирането на обектите. Класификаторите формират "общия език", който позволява данните да се групират по значение и данните да се интегрират между различни системи, нива на управление и фази от жизнения цикъл на проекта.

Най-осезаемото въздействие на класификаторите е в икономиката на жизнения цикъл на сградата, където най-важният аспект е оптимизирането на дългосрочните експлоатационни разходи. Проучванията показват, че експлоатационните разходи съставляват до 80% от общите разходи за притежание на сградата, което е три пъти повече от първоначалните разходи за строителство (фиг. 4.2-6) [98]. Това означава, че решението за бъдещите разходи до голяма степен се формира на етапа на проектиране

Ето защо изискванията от оперативните инженери (CAFM, AMS, PMS, RPM) трябва да станат отправна точка за генериране на изисквания към данните по време на фазата на проектиране (фиг. 1.2-4). Тези системи не трябва да се разглеждат като краен етап на проекта, а като неразделна част от цялата цифрова екосистема на проекта - от концепцията до демонтажа

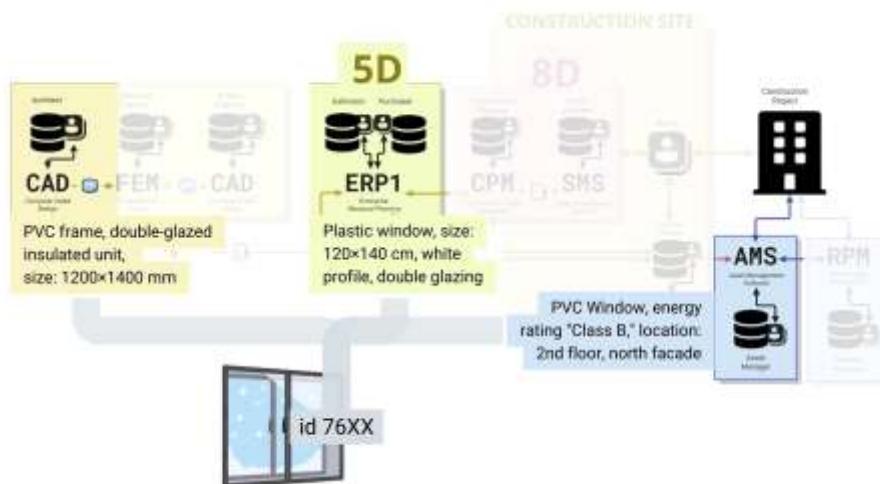
Модерен класификатор не е просто система от кодове за групиране. Той е механизъм за взаимно разбиране между архитекти, инженери, оценители, логистици, поддръжка и ИТ системи. Точно както автопилотът на автомобила трябва да разпознава недвусмислено и точно пътните обекти, така и системите за цифрово строителство и техните потребители трябва да интерпретират един и същ елемент на проекта недвусмислено за различните системи чрез класа на элемента.



**Фиг. 4.2-6** Разходите за експлоатация и поддръжка надвишават три пъти разходите за строителство, като съставляват 60-80% от общите разходи за жизнения цикъл на сградата (според [99]).

Нивото на развитие на класификатора е пряко свързано с дълбочината на цифровизацията на компанията и нейната цифрова зрялост. Организациите с ниско ниво на цифрова зрялост се сблъскват с фрагментирани данни, несъвместими информационни системи и в резултат на това с несъвместими и неефективни класификатори. В такива компании един и същ елемент често може да има различни идентификатори за групиране в различните системи, което критично затруднява крайната интеграция и прави автоматизацията на процесите невъзможна.

Например един и същ прозорец в проекта може да бъде обозначен по различен начин в CAD модел, система за оценяване и поддръжка (фиг. 4.2-7) поради многоизмерното възприемане на елементите от различните участници в процеса. За оценителя в категорията прозорци са важни елементът, обемът и цената, за службата за поддръжка - наличността и възможността за поддръжка, за архитекта - естетическите и функционалните характеристики. В резултат на това един и същ елемент може да изисква различни параметри.



**Фиг. 4.2-7** При непоследователна класификация между системите, елементът ще губи част от информацията за атрибута на всяка стъпка от прехода си към друга система.

Поради трудността да се определи еднозначно класификацията на строителните елементи, специалисти от различни области често определят несъвместими класове за един и същ елемент. Това води до загуба на единна представа за обекта, което изисква последваща ръчна намеса за хармонизиране на различните системи за класификация и за установяване на съгласуваност между типовете и класовете, определени от различните специалисти.

В резултат на това несъответствие оперативната документация, получавана от отдела за обществени поръчки (ERP) при закупуването на строителен елемент от производителя, често не може да бъде правилно свързана с класификацията на този елемент на строителната площадка (PMIS, SCM). Вследствие на това е малко вероятно критичната информация да бъде интегрирана в системите за управление на инфраструктурата и активите (CAFM, AMS), което създава сериозни проблеми по време на въвеждането в експлоатация, както и при последващата поддръжка (AMS, RPM) или подмяна на элемента.

В компаниите с висока степен на цифрова зрялост класификаторите играят ролята на нервна система, която обединява всички информационни потоци. Един и същ елемент получава уникален идентификатор, който му позволява да се прехвърля между CAD, ERP, AMS и CAFM -системите и техните класификатори без изкривяване или загуба.

За да създадете ефективни класификатори, трябва да разберете как се използват данните. Един и същ инженер може да назове и класифицира даден елемент по различен начин в различните проекти. Само чрез събиране на статистически данни за използването през годините може да се разработи стабилна система за класификация. За това помага машинното обучение: алгоритми анализират хиляди проекти (фиг. 9.1-10), като определят вероятни класове и параметри чрез машинно обучение (фиг. 10.1-6). Автоматичната класификация е особено ценна в среди, в които ръчната класификация не е възможна поради обема на данните. Системите за автоматична класификация ще могат да разграничават основни категории въз основа на минимално попълнени параметри на елементите (повече подробности в девета и десета част на книгата).

Разработените системи за класификация се превръщат в катализатор за по-нататъшна цифровизация, създавайки основа за:

- Автоматизирано оценяване на разходите и сроковете на проекта.
- Прогнозен анализ на потенциални рискове и конфликти
- Оптимизиране на процесите на снабдяване и логистичните вериги
- Създаване на цифрови близнаци на сгради и структури
- Интеграция със системи за интелигентен град и интернет на нещата

Времето за трансформация е ограничено - с развитието на технологиите за машинно обучение и компютърно зрение проблемът с автоматичната класификация, който от десетилетия е нерешим, ще бъде решен през следващите години, а строителните и проектантските компании, които не успеят да се адаптират навреме, рискуват да повторят съдбата на таксиметровите автопаркове, изместени от цифровите платформи.

Автоматизацията на изчисленията на разходите и графиците, както и големите данни и машинното

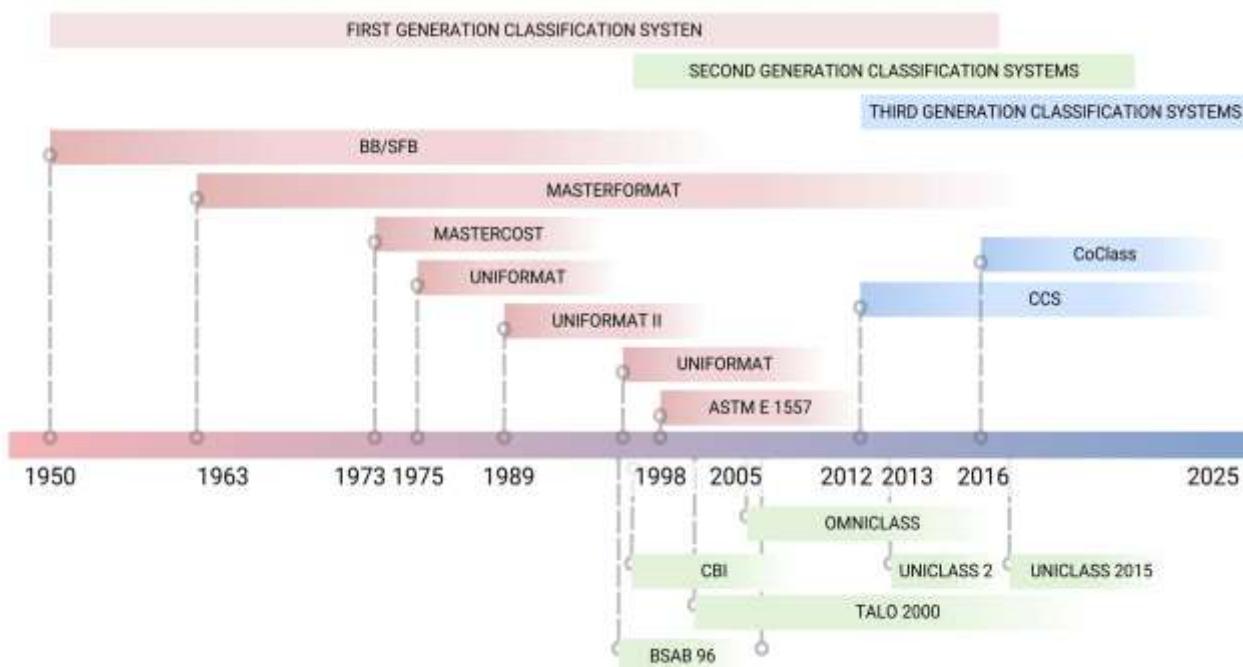
обучение, ще бъдат разгледани по-подробно в пета и девета част на книгата. Рискът от повторение на съдбата на таксиметровите автопаркове и Uberизацията на строителния бранш са разгледани подробно в десетата част на книгата.

За да се разбере ключовата роля на класификаторите в цифровата трансформация на строителната индустрия, е необходимо да се обрънем към историята на тяхното развитие. Именно историческият контекст ни позволява да осъзнаем как са се развили подходите към класификацията и какви тенденции определят сегашното им състояние.

## Masterformat, OmniClass, Uniclass и CoClass: еволюция на системите за класификация

Исторически погледнато, класификаторите на строителни елементи и работи са се развивали в три поколения, всяко от които е отразявало нивото на наличните технологии и текущите нужди на индустрията в определен период от време (фигура 4.2-8):

- **Първо поколение** (от началото на 50-те до края на 80-те години на XX век) - хартиени указатели, йерархични класификатори, използвани на местно ниво (напр. Masterformat, SfB).
- **Второто поколение** (от края на 90-те до средата на 2010 г.) са електронни таблици и структурирани бази данни, реализирани в Excel и Access (ASTM E 1557, OmniClass, Uniclass 1997).
- **Трето поколение** (от 2010 г. до сега) - цифрови услуги и API-интерфейси, интеграция с CAD (BIM), автоматизация (Uniclass 2015, CoClass).



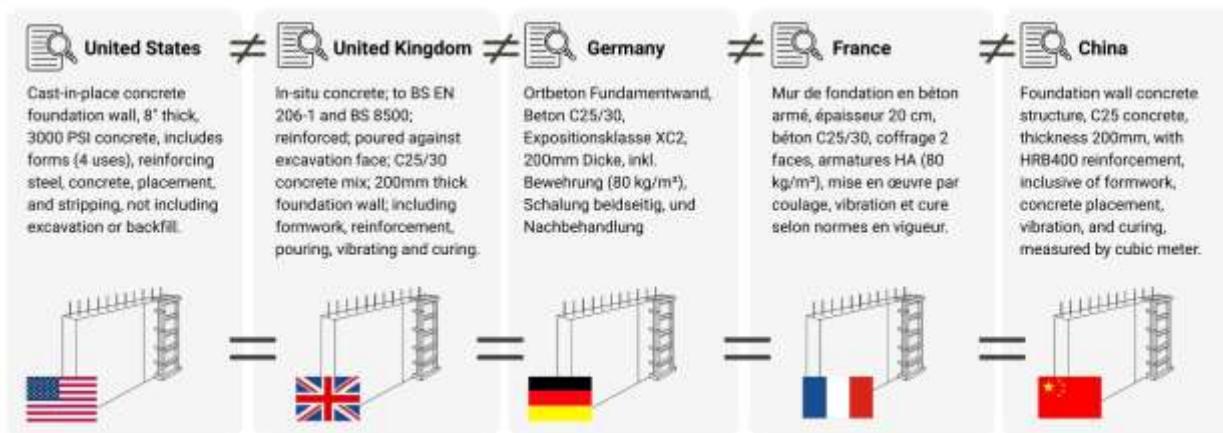
Фиг. 4.2-8 Три поколения класификатори за строителната индустрия.

През последните десетилетия се наблюдава намаляване на йерархичната сложност (фиг. 4.2-9) на класификаторите: докато ранните системи, като OmniClass, използваха до 7 нива на влагане за описание на 6887 класа, съвременните решения, като CoClass, са ограничени до 3 нива със 750 класа. Това улеснява работата с данните, като същевременно се запазва необходимата детайлност. Uniclass 2015, който често се използва като стандарт в Обединеното кралство, обединява 7210 класа в само 4 нива, което го прави удобен за проекти на CAD и обществени поръчки.

Classifier	Table / Objects	Number of classes	Nesting depth
OmniClass	Table 23 Products	6887	7 levels
Uniclass 2015	Pr — Products	7210	4 levels
CoClass, CCS	Components	750	3 levels

Фиг. 4.2-9 С всяко ново поколение класификатори сложността на категоризацията намалява скокообразно.

В системите за оценка на строителството в различните страни, поради различията в класификациите, дори един типичен елемент като бетонна стена за основи може да бъде описан по много различни начини (фигура 4.2-10). Тези разлики отразяват националните строителни практики, използваните системи за измерване, подходите към класификацията на материалите, както и нормативните и техническите изисквания, действащи във всяка страна.



Фиг. 4.2-10 Един и същ елемент се използва в проекти в различни държави чрез различни описания и класификации.

Разнообразието от класификации на едни и същи елементи усложнява международното сътрудничество и прави сравнението на разходите и обхвата на работата в рамките на международни проекти времеемко, а понякога почти невъзможно. Понастоящем не съществува универсален класификатор на световно ниво - всяка страна или регион разработва свои собствени системи, основани на местните норми, език и бизнес култура:

- **CCS** (Дания): Система за класифициране на разходите - система за класифициране на разходите през целия жизнен цикъл на съоръжението (проектиране, изграждане, експлоатация). Акцентът е върху логиката на експлоатацията и поддръжката, но включва и управлението на бюджета и ресурсите.
- **NS 3451** (Норвегия): категоризира съоръженията по функции, елементи на дизайна и етапи на жизнения цикъл. Използва се за управление на проекти, оценка на разходите и дългосрочно планиране.
- **MasterFormat** (САЩ): система за структуриране на строителните спецификации в раздели (напр. бетон, електричество, довършителни работи). Фокус върху дисциплини и видове работи, а не върху функционални елементи (за разлика от UniFormat).
- **Uniclass 2** (Великобритания): един от най-подробните класификатори, използвани в проектите за обществени поръчки и BIM. Обединява данните за обекти, работи, материали и пространства в единна система.
- **OmniClass**: международен стандарт (разработен от CSI в САЩ) за управление на информация за обекти от библиотеки с компоненти до електронни спецификации. Подходящ за дългосрочно съхранение на данни, съвместим с CAD (BIM) и други цифрови инструменти.
- **COBie**: Обмен на информация за строителството и експлоатацията на сгради е международен стандарт за обмен на данни между фазите на проектиране, строителство и експлоатация. Включен е в BS 1192-4:2014 като част от концепцията "BIM -модел, готов за използване". Фокусира се върху трансфера на информация (напр. спецификации на оборудването, гаранции, контакти с изпълнители).

Глобализацията на строителната индустрия вероятно ще доведе до постепенно унифициране на системите за класификация на строителните елементи, което значително ще намали зависимостта от местните национални стандарти. Този процес може да се развие по подобие на еволюцията на интернет комуникациите, където универсалните протоколи за трансфер на данни в крайна сметка изместиха различните местни формати, осигурявайки глобална оперативна съвместимост на системите.

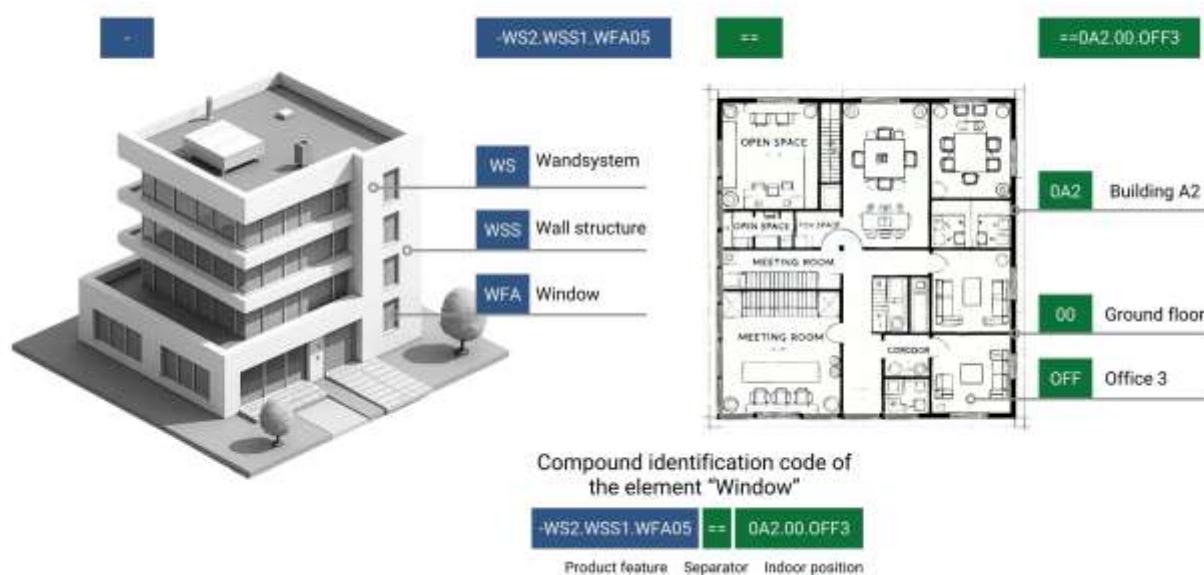
Алтернативен път за развитие може да бъде директният переход към системи за автоматична класификация, основани на технологии за машинно обучение. Тези технологии, които днес се разработват основно в областта на автономния транспорт, имат значителен потенциал за приложение към големи масиви от данни за CAD проектиране (фиг. 10.1-6).

Днес ситуацията не се ограничава само до националното групиране на класификатори. Поради многото особености, които не се отчитат на национално ниво, всяко предприятие трябва да унифицира и стандартизира категориите елементи и ресурси, с които работи.

По правило този процес започва с малки обекти - локални таблици на обектите или вътрешни системи за етикетиране. Стратегическата цел обаче е да се премине към общ език за описание на всички елементи, който да е разбираем не само в рамките на компанията, но и извън нея - в идеалния случай хармонизиран с международни или отраслови класификатори (фиг. 4.2-8). Този подход улеснява интеграцията с външни партньори, цифрови системи и насярчава формирането на унифицирани процеси от край до край в рамките на жизнения цикъл на обектите.

Преди да се премине към автоматизация и мащабираме ИТ системи, е необходимо да се

използват класификатори на национално ниво или да се изгради собствена, логична и недвусмислена структура за идентификация на елементите. Всеки обект - независимо дали е прозорец (фигура 4.2-11), врата или инженерна система - трябва да бъде описан по такъв начин, че да може да бъде разпознат безпогрешно във всяка цифрова система на компанията. Това е от решаващо значение при прехода от плоски чертежи към цифрови модели, обхващащи както фазата на проектиране, така и експлоатацията на сградата.



**Фиг. 4.2-11 Пример за съставен идентификатор на прозоречен елемент на сградата въз основа на класификация и позиция в сградата.**

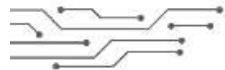
Един пример за вътрешни класификатори може да бъде разработването на съставен идентификационен код (фигура 4.2-11). Такъв код съчетава няколко нива на информация: функционалното предназначение на елемента (например "прозорец в стената"), неговия тип и точната пространствена референция - сграда А2, етаж 0, стая 3. Такава многостепенна структура позволява да се създаде единна система за навигация из цифровите модели и документация, особено на етапите на проверка и преобразуване на данните, където се изисква еднозначно групиране на елементите. Еднозначното разпознаване на елементите осигурява съгласуваност между отделите и намалява рисковете от дублиране, грешки и загуба на информация.

Добре изграденият класификатор не е просто технически документ, той е основата на цифровата екосистема на компанията:

- осигурява съвместимост на данните между системите;
- намалява разходите за търсене и обработка на информация;
- увеличава прозрачността и управляемостта;
- създава основа за мащабиране и автоматизация.

Стандартизираното описание на обектите чрез използване на национални класификатори или патентовани съставни идентификационни кодове се превръща в основа за последователни данни, надежден обмен на информация и последващо прилагане на интелигентни услуги - от автоматизирано възлагане на обществени поръчки до цифрови близнаци.

След приключване на етапа на структуриране на многоформатните данни и избора на класификатор, който ще се използва за разпознаване и групиране на елементите, следващата стъпка е правилното моделиране на данните. Този процес включва идентифициране на ключови параметри, изграждане на логическа структура на данните и описание на връзките между елементите.



## ГЛАВА 4.3.

# МОДЕЛИРАНЕ НА ДАННИ И ЦЕНТЪР ЗА ВЪРХОВИ ПОСТИЖЕНИЯ

### Моделиране на данни: концептуален, логически и физически модел

Ефективното управление на данни (структурирани и категоризирани от нас по-рано) е невъзможно без добре обмислена структура за съхранение и обработка. За да осигурят достъп и последователност на информацията на етапите на съхранение и обработка, компаниите използват моделиране на данни - методология, която им позволява да проектират таблици, бази данни и връзки между тях в съответствие с бизнес изискванията.

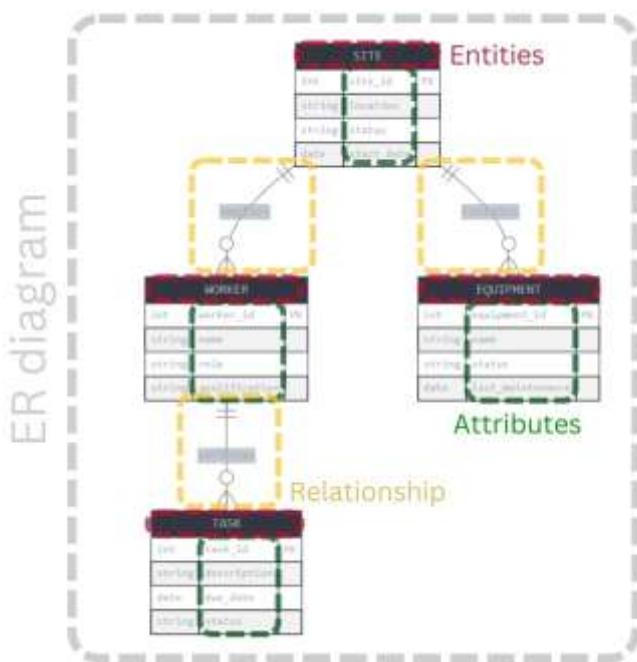
Моделирането на данни е основата, върху която се изгражда всяка цифрова екосистема. Без описание на системите, изискванията и моделирането на данни инженерите и специалистите, които създават данни, не знаят и не разбират къде ще бъдат използвани създадените от тях данни.

Подобно на строежа на сграда, при който не можете да започнете да слагате тухли без план, създаването на система за съхранение на данни изисква ясно разбиране за това какви данни ще се използват, как ще бъдат свързани и кой ще работи с тях. Без описание на процесите и изискванията инженерите и специалистите, създаващи данните, губят представа къде и как ще бъдат използвани данните в бъдеще.

Моделът на данните служи като мост между бизнеса и ИТ. Той позволява да се формализират изискванията, да се структурира информацията и да се улесни комуникацията между заинтересованите страни. В този смисъл моделирането на данни прилича на работата на архитекта, който според плана на клиента разработва план на сградата и след това го предава на строителите - администратори и разработчици на бази данни - за изпълнение (създаване на база данни).

Така всяка строителна компания, освен че трябва да структурира и категоризира елементите и ресурсите (фиг. 4.2-11), трябва да овладее изкуството да "строи" бази данни (таблици) и да се научи да създава връзки между тях, сякаш свързвайки тухлите в надеждна и здрава стена от знания от данните на компанията. Ключовите понятия в моделирането на данни (фигура 4.3-1) включват:

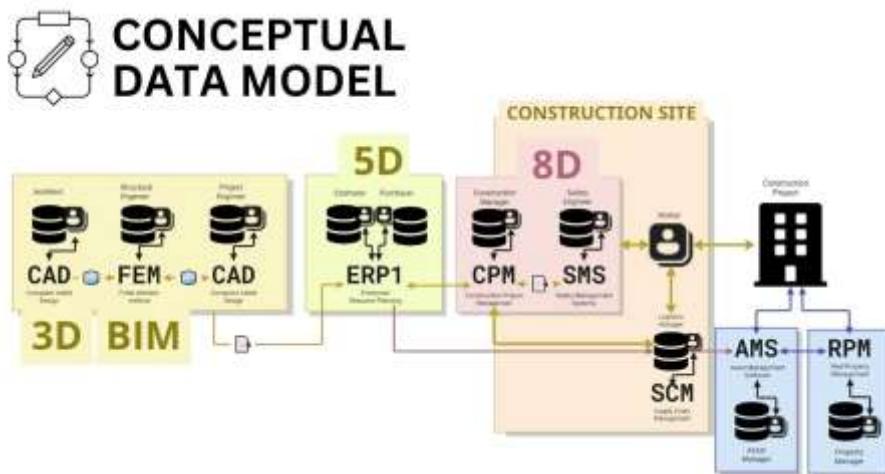
- **Субектите** са обектите, за които трябва да се събират данни. В ранната фаза на проектиране единицата може да бъде единичен елемент (напр. "врата"), а в модела за оценяване - група от категоризирани елементи (напр. "вътрешни врати").
- **Атриутите** са характеристики на обекти, които описват важни подробности: размери, свойства, разходи за сглобяване, логистика и други параметри.
- **Връзки (линкове)** - показват как същностите си взаимодействват една с друга. Те могат да бъдат от един от видовете: "едно към едно", "много към едно", "много към много".
- **ER диаграмите (Entity-Relationship diagrams)** са визуални диаграми, които показват същности, атрибути и връзките между тях. ER-диаграмите могат да бъдат концептуални, логически и физически - всяка от тях отразява различно ниво на детайлност.



Фигура 4.3-1 ER диаграма на концептуална структура на база данни със същности, атрибути и връзки.

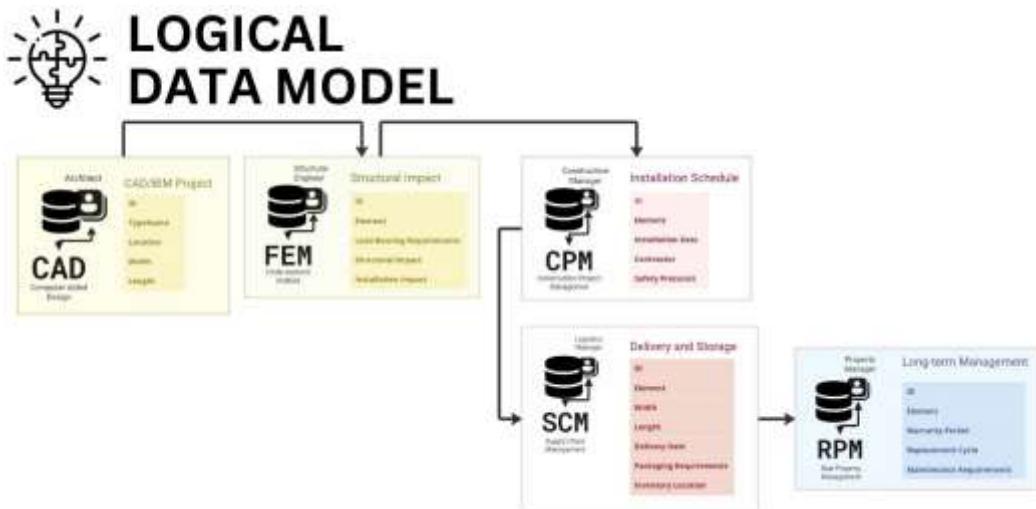
Процесът на проектиране на данни и определяне на връзките между тях традиционно се разделя на три основни модела. Всеки от тях изпълнява определени функции, като се различава по нивото на детайлност и степента на абстракция при представянето на структурата на данните:

- **Концептуален модел на данните:** този модел описва основните същности и техните взаимоотношения, без да навлиза в подробности за атрибутите. Обикновено се използва в началните етапи на планирането. На този етап можем да направим скица от базите данни и системите, за да покажем връзката между различните отдели и специалисти.



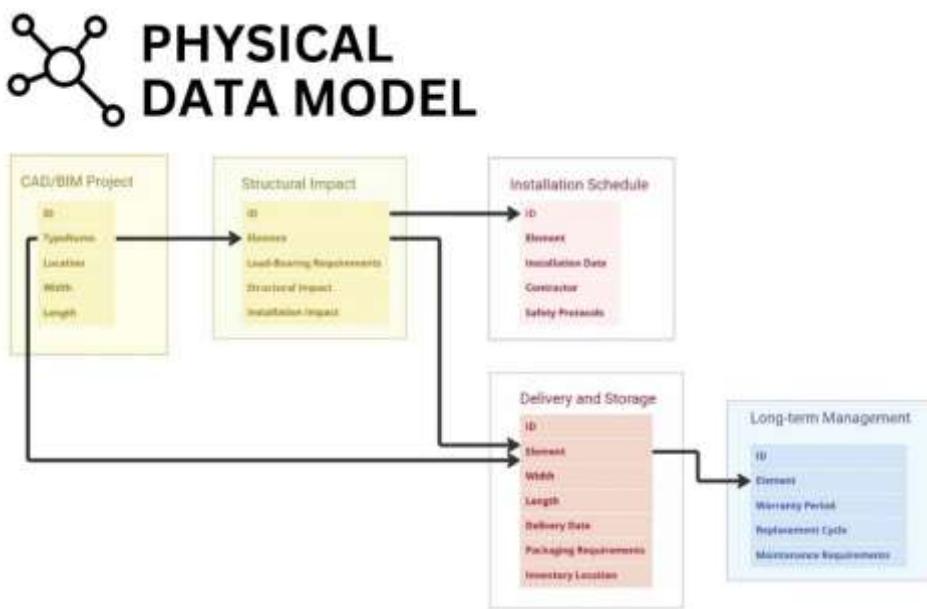
Фиг. 4.3-2 Концептуалната диаграма описва съдържанието на системата: представяне на връзките на високо ниво, без технически подробности.

- Логически модел на данните:** Въз основа на концептуалния модел, логическият модел на данните включва подробни описания на същности, атрибути, ключове и връзки, картографиране на бизнес информация и правила.



Фигура 4.3-3 Логически модел на данните описва подробно типовете данни, връзките и ключовете, но без системна реализация.

- Физически модел на данните:** Този модел описва необходимите структури за реализиране на база данни, включително таблици, колони и връзки. Той се фокусира върху производителността на базите данни, стратегиите за индексиране и физическото съхранение, за да се оптимизира физическото разгръщане на базите данни.

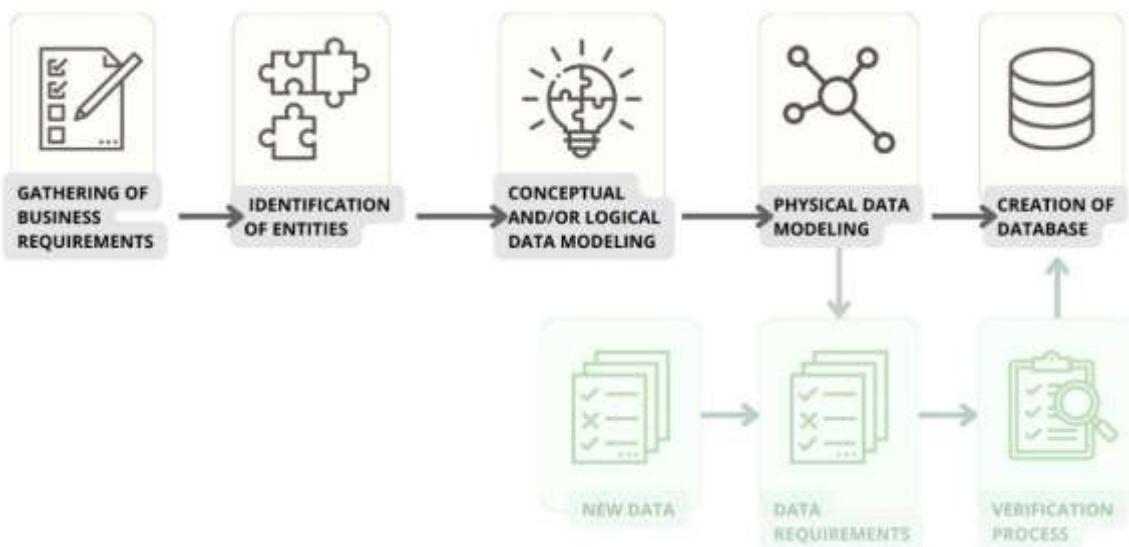


Фигура 4.3-4 Физическият модел на данните определя начина, по който ще бъде реализирана системата, включително табличите и специфичните детайли на базата данни.

При проектирането на бази данни и на таблични връзки разбирането на нивата на абстракция играе ключова роля за изграждането на ефективна системна архитектура.

Ефективната методология за моделиране на данни ви позволява да съчетаете бизнес целите с техническото изпълнение, което прави цялата верига от процеси по-прозрачна и управляема. Моделирането на данни не е еднократна задача, а процес, включващ последователни стъпки (фигура 4.3-5):

- **Събиране на бизнес изисквания:** дефинират се ключови задачи, цели и информационни потоци. Това е етапът на активно взаимодействие с експерти и потребители.
- **Идентифициране на субектите:** подчертават се основните обекти, категории и типове данни, които е важно да се вземат предвид в бъдещата система.
- **Разработване на концептуален и логически модел:** първо се улавят ключовите същности и техните взаимоотношения, а след това атрибутите, правилата и подробната структура.
- **Физическо моделиране:** проектира се техническата реализация на модела: таблици, полета, връзки, ограничения, индекси.
- **Създаване на база данни:** последната стъпка е да се реализира физическият модел в избраната СУБД, да се извърши тестване и да се подготви за работа.



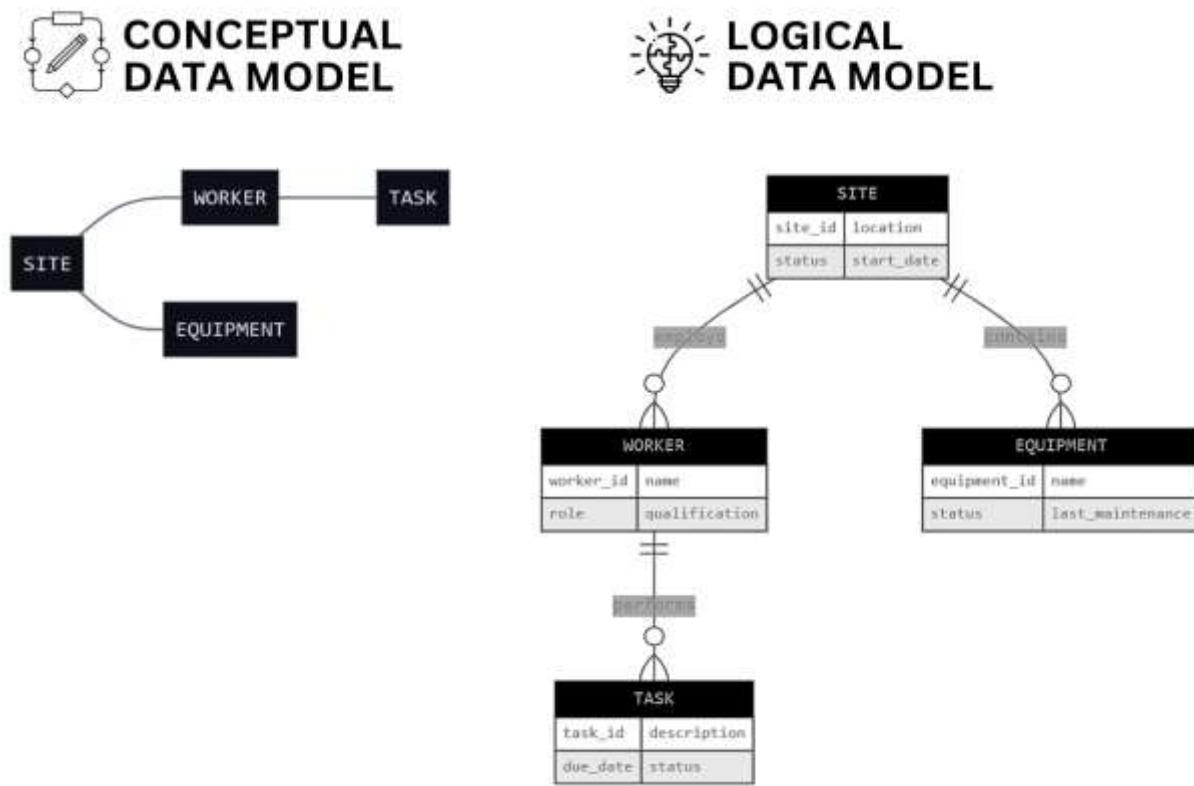
**Фиг. 4.3-5 Създаването на бази данни и системи за управление на данни за бизнес процеси започва с генериране на изисквания и моделиране на данни.**

Правилно разработените процеси за моделиране на данни позволяват прозрачни информационни потоци, което е особено важно при сложни проекти, като например управление на строителни проекти или обекти. Нека разгледаме как преминаването от концептуален модел към логически модел и след това към физически модел може да помогне за рационализиране на процесите.

## Практическо моделиране на данни в контекста на строителството

Нека да вземем за пример за моделиране на данни задача за управление на строителен обект и да преобразуваме изискванията на бригадира в структуриран логически модел. Въз основа на основните нужди на управлението на строителната площадка дефинираме ключови същности за: площадка (SITE), работници (WORKER), оборудване (EQUIPMENT), задачи (TASK) и използване на оборудването (EQUIPMENT\_USAGE). Всяка същност съдържа набор от атрибути, които отразяват важни характеристики. Например за TASK (Задача) това може да бъде описание на задачата, дата на изпълнение, статус, приоритет; за WORKER (Работник) това може да бъде името, ролята му на обекта, настоящата му заетост и т.н.

Логическият модел установява връзките между тези единици, като показва как те си взаимодействат в реалните работни процеси (фигура 4.3-6). Например връзката между обект и работници показва, че много работници могат да работят на един обект, докато връзката между работници и задачи отразява, че един работник може да изпълнява множество задачи.

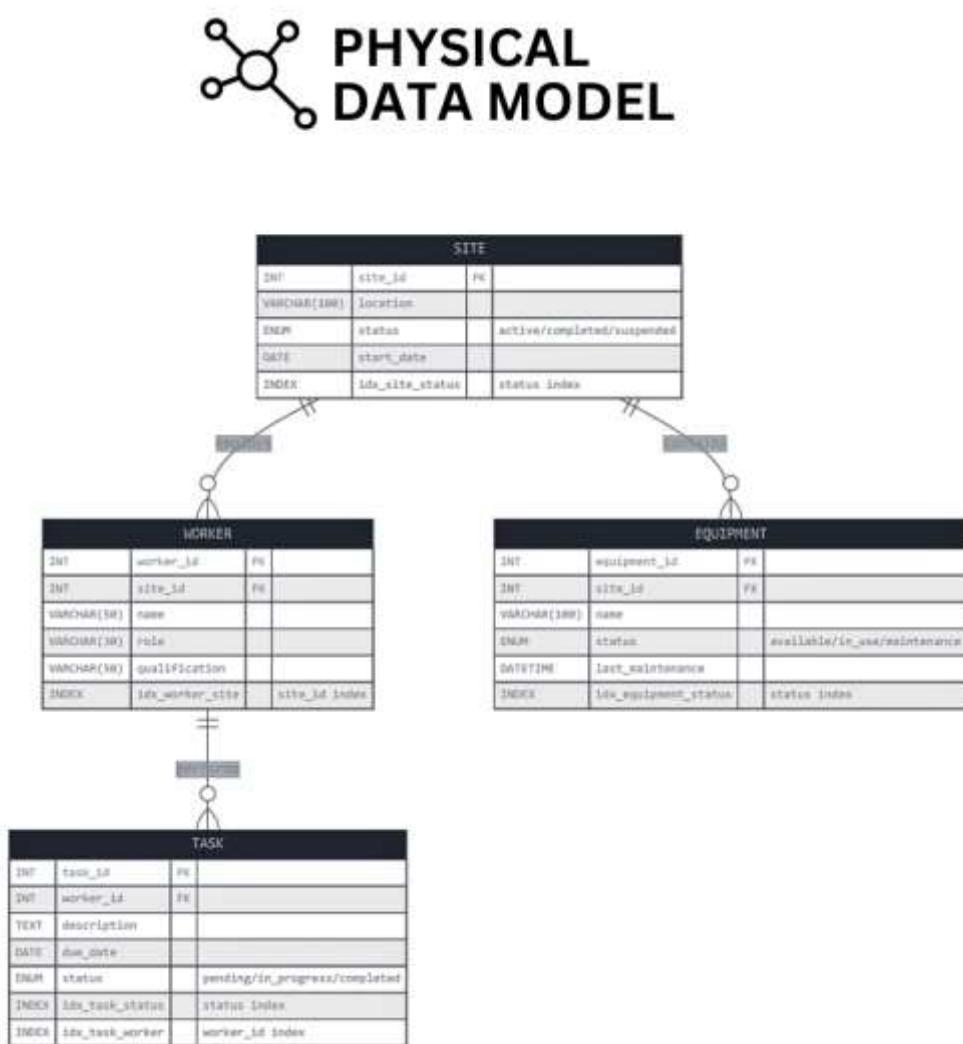


Фиг. 4.3-6 Концептуален и логически модел на данните, генериран от изискванията на бригадира за описание на процесите на строителната площадка.

Когато се премине към физическия модел, се добавят технически подробности за изпълнението: специфични типове данни (VARCHAR, INT, DATE), първични и външни ключове за връзките между таблиците и индекси за оптимизиране на производителността на базата данни (фигура 4.3-7).

Например за статусите трябва да се определят специфични типове с възможни стойности и да се добавят индекси на ключови полета като status и worker\_id, за да се подобри ефективността на търсачката. По този начин логическото описание на системата се превръща в конкретен план за

реализация на базата данни, готов за създаване и внедряване.



Фигура 4.3-7 Моделът на физическите данни описва същностите на дадена строителна площадка чрез минимално необходимите параметри.

Физическият модел често се различава от логическия модел. Средното разпределение на времето за моделиране е следното: около 50% се изразходват за концептуалния модел (събиране на изисквания, обсъждане на процеси, идентифициране на същности), 10% за логическия модел (определяне на атрибути и връзки) и 40% за физическия модел (внедряване, тестване, адаптиране към СУБД).

Това равновесие се дължи на факта, че концептуалният етап полага основите на структурата на данните, докато логическият модел само определя връзките и атрибутите. Физическият модел изиска най-много ресурси, тъй като именно на този етап данните се внедряват в конкретни платформи и инструменти.

## Създаване на база данни с помощта на LLM

След като имаме модел на данните и описание на същностите чрез параметри, сме готови да създадем бази данни - хранилища, в които ще съхраняваме информацията, постъпила след етапа на структуриране на конкретни процеси.

Нека се опитаме да създадем пример за проста, но функционална база данни с минимално количество код, като използваме SQLite на примера на езика за програмиране Python. Релационните бази данни бяха разгледани подробно в главата "Структурирани релационни бази данни и език за заявки SQL".

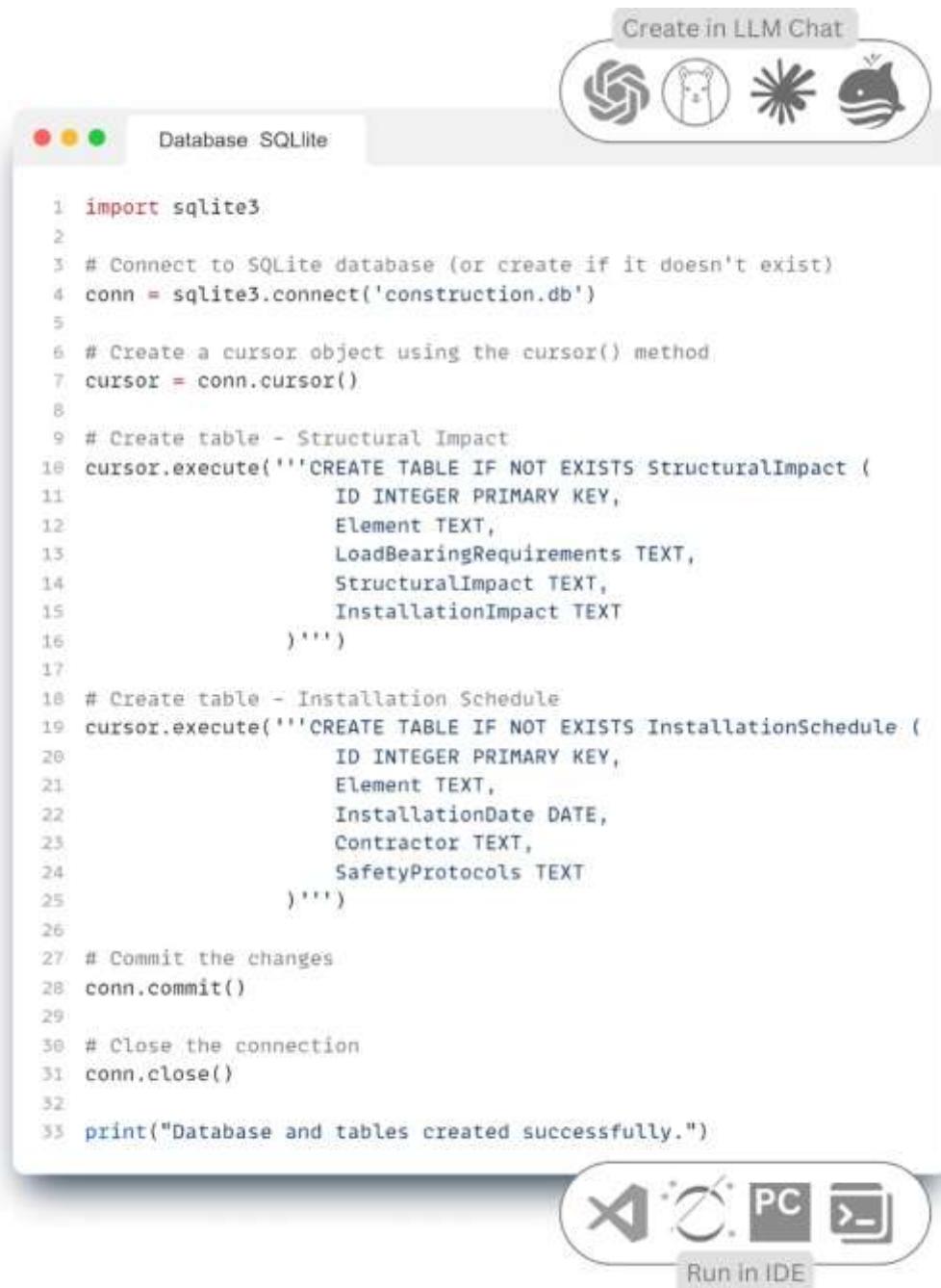
SQLite е олекотена, вграждаща се, междуплатформена база данни с отворен код, поддържаща основните функции на SQL. SQLite се използва широко в CAD (BIM), ERP, EPM и присъства в почти всеки смартфон.

Нека да поискаме от модела LLM да създаде за нас малка база данни SQLite въз основа на физическия модел, създаден по-горе (фигура 4.3-4) за системите FEM (структурно въздействие) и CPM (график на сглобяване).

- ❷ Из pratете текстова заявка до чата LLM (ChatGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN или някой друг):

Създайте SQLite бази данни за системите "Конструктивно въздействие" и "График за инсталiranе" от фигурата в прикачения файл (поставете схемата в чата като изображение (фиг. 4.3-4) ↴

❑ Отговор на LLM :



The screenshot shows a window titled "Database SQLite". At the top right is a button labeled "Create in LLM Chat" with four icons: a neural network, a bear, a flower, and a whale. Below the title bar is a toolbar with three colored circles (red, yellow, green). The main area contains the following Python code:

```

1 import sqlite3
2
3 # Connect to SQLite database (or create if it doesn't exist)
4 conn = sqlite3.connect('construction.db')
5
6 # Create a cursor object using the cursor() method
7 cursor = conn.cursor()
8
9 # Create table - Structural Impact
10 cursor.execute('''CREATE TABLE IF NOT EXISTS StructuralImpact (
11         ID INTEGER PRIMARY KEY,
12         Element TEXT,
13         LoadBearingRequirements TEXT,
14         StructuralImpact TEXT,
15         InstallationImpact TEXT
16     )''' )
17
18 # Create table - Installation Schedule
19 cursor.execute('''CREATE TABLE IF NOT EXISTS InstallationSchedule (
20         ID INTEGER PRIMARY KEY,
21         Element TEXT,
22         InstallationDate DATE,
23         Contractor TEXT,
24         SafetyProtocols TEXT
25     )''' )
26
27 # Commit the changes
28 conn.commit()
29
30 # Close the connection
31 conn.close()
32
33 print("Database and tables created successfully.")

```

At the bottom right is a button labeled "Run in IDE" with icons for a terminal, a PC, and a file.

**Фигура 4.3-8 Генерираният код създава бази данни и таблици от каченото изображение, в които сега можем да заредим данни.**

Този код за база данни може да се изпълнява в средата на Python с поддръжка на SQLite, която обикновено е включена в стандартната среда на Python. Той може да бъде изпълнен на всяка локална Python машина, способна да изпълнява скриптове и да създава файлове. След като бъде стартиран и изпълнен, този код ще създаде файл на вашия твърд диск - SQLite база данни с име **construction.db** (подобно на Excel файл с работни листове), който ще съдържа таблиците "Структурно въздействие" и "График на монтажа".

След като бъде създадена нова база данни **construction.db** в тези таблици, към тях могат да се добавят данни чрез SQL \_запитвания или импортиране, което ще ви позволи да започнете да създавате автоматична обработка на данни по-късно. Данните могат да се импортират в SQLite базата данни от CSV файлове, електронни таблици в Excel или да се експортират от други бази данни и хранилища чрез API.

За да създаде устойчиви процеси за моделиране на данни и ефективно управление на бази данни, компанията се нуждае от ясно дефинирана стратегия, както и от координация между техническите и бизнес екипите. При разнородни проекти и множество източници на данни често е трудно да се осигури последователност, стандартизация и контрол на качеството на всички нива. Едно от ключовите решения може да бъде създаването на специален център за върхови постижения в областта на моделирането на данни (CoE) в рамките на компанията.

## **Център за върхови постижения (ЦВП) за моделиране на данни**

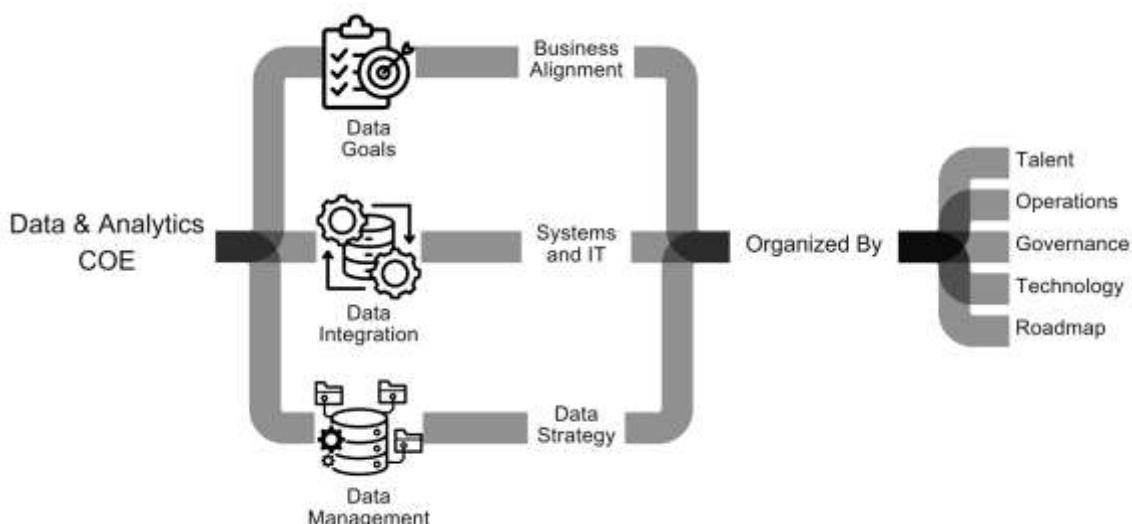
Тъй като данните се превръщат в един от ключовите стратегически активи, компаниите трябва да правят нещо повече от това да събират и съхраняват информацията правилно - важно е да се научат как да управляват данните систематично. Центърът за върхови постижения за класификация и моделиране на данни (CoE) е структурно звено, което осигурява последователност, качество и ефективност на цялата работа с данни в организацията.

Центрът за върхови постижения (ЦВП) е ядрото на експертната подкрепа и методологическата основа за цифровата трансформация в компанията. Той изгражда култура, базирана на данни, и дава възможност на организациите да изграждат процеси, които вземат решения въз основа на структурирани, валидирани и представителни данни, а не на интуиция или местна информация.

Центрът за върхови постижения в областта на данните обикновено се формира от многофункционални екипи, които работят на принципа "две пици". Този принцип, предложен от Джейф Безос, означава, че размерът на екипа трябва да е такъв, че да може да се нахрани с две пици, т.е. не повече от 6-10 души. Този подход помага да се избегне прекомерната бюрокрация и увеличава гъвкавостта на работата. Екипът на CoE трябва да включва служители с различни технически умения - от анализа на данни и машинното обучение до експертни познания в конкретни бизнес области. Със своите задълбочени технически познания инженерите по данни следва не само да оптимизират процесите и да моделират данните, но и да подпомагат колегите си, като намаляват времето за изпълнение на рутинни задачи (фиг. 4.3-9).

Както в природата устойчивостта на екосистемите се осигурява от биоразнообразието, така и в цифровия свят гъвкавостта и адаптивността се постигат чрез разнообразни подходи за обработка на данни. Това разнообразие обаче трябва да бъде подкрепено от общи правила и концепции.

Центърът за върхови постижения (СоЕ) може да бъде сравнен с "климатичните условия" на една горска екосистема, които определят кои видове данни ще процъфтяват и кои ще бъдат автоматично отхвърлени. Създавайки благоприятен "климат" за качествени данни, СЕ улеснява естествения подбор на най-добрите практики и методологии, които след това се превръщат в стандарти за организацията.



**Фиг. 4.3-9 Центърът за върхови постижения (СоЕ) за данни и анализи обединява експертни познания по ключови аспекти на управлението, интеграцията и стратегията на данните.**

За да се ускорят циклите на интеграция и да се постигнат по-добри резултати, СЕ трябва да предложи на членовете си достатъчна степен на автономност при вземането на решения. Това е особено важно в динамична среда, където пробите и грешките, постоянната обратна връзка и честите пускания на нови версии могат да донесат значителни ползи. Тази автономия обаче е ефективна само при наличието на ясна комуникация и подкрепа от страна на висшето ръководство. Без стратегическа визия и координация на най-високо ниво дори и най-компетентният екип може да се сблъска с пречки при изпълнението на своите инициативи.

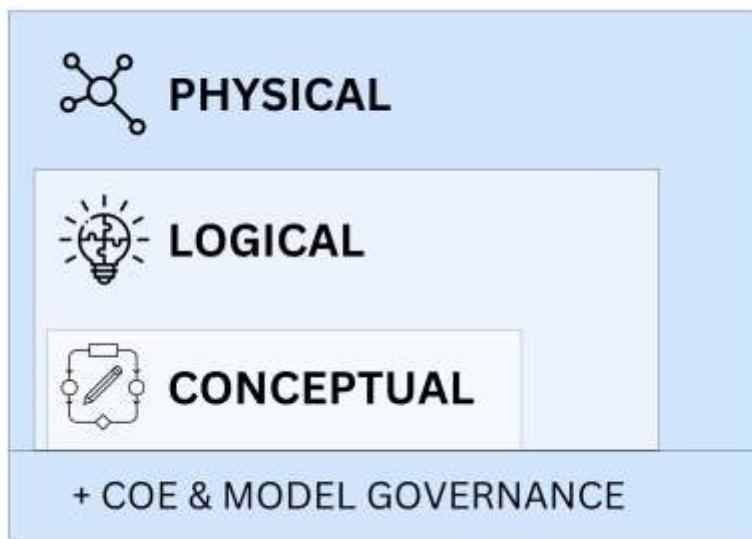
СОЕ или висшето ръководство на компанията са тези, които отговарят за това подходът за моделиране на данни да не се ограничава само до един или два проекта, а да бъде вграден в цялостната система за управление на информацията и бизнес процесите.

Центърът за експертни познания (ЦЕ), в допълнение към задачите си, свързани с моделирането на данни и управлението на данни, отговаря за разработването на общи стандарти и подходи за

внедряване и експлоатация на инфраструктурата за данни. Освен това той насърчава културата на непрекъснато усъвършенстване, оптимизиране на процесите и ефективно използване на данните в организацията (фигура 4.3-10).

Систематичният подход към управлението на данни и модели в рамките на СоЕ може да бъде разделен на няколко основни блока:

- **Стандартизиране на процесите и управление на жизнения цикъл на моделите:** СоЕ разработва и прилага методологии за унифициране на създаването и управлението на моделите на данни. Това включва: създаване на структурни шаблони, методи за контрол на качеството и системи за контрол на версии, за да се осигури непрекъснатост на данните във всички фази на работа.
- **Управление на ролите и разпределение на отговорностите:** СоЕ определя ключовите роли в процеса на моделиране на данни. На всеки участник в проекта се възлагат ясно определени роли и области на отговорност, което улеснява работата в екип и намалява риска от несъответствия в данните.
- **Контрол на качеството и одит:** ефективното управление на данните за строителството изиска непрекъснато наблюдение на тяхното качество. Въвеждат се автоматизирани механизми за проверка на данните, идентифициране на грешки и липсващи атрибути.
- **Управление на метаданни и информационна архитектура:** СоЕ отговаря за създаването на единна система за класификация и идентификатори, стандарти за именуване и описание на обекти, което е от решаващо значение за интеграцията между системите.



**Фиг. 4.3-10 Моделирането на данни и управлението на качеството на данните е едно от основните предизвикателства пред СоЕ**

Центрърът за върхови постижения (СоЕ) за данни не е просто група от експерти, а системен механизъм, който създава нова култура, основана на данни, и осигурява единен подход за работа с данни в цялата компания. Чрез компетентно интегриране на процесите на моделиране в цялостната система за управление на информацията, стандартизация, класификация и контрол на качеството на данните, СоЕ помага на предприятията непрекъснато да подобряват своите продукти и бизнес процеси, да реагират по-бързо на промените на пазара и да вземат информирани решения

въз основа на надеждни анализи.

Такива центрове са особено ефективни, когато се комбинират със съвременните принципи на DataOps - под ход, който осигурява непрекъснато предоставяне, автоматизация и контрол на качеството на данните. Ще говорим повече за DataOps в Част 8, в главата "Съвременни технологии за данни в строителната индустрия".

В следващите глави ще преминем от стратегията към практиката - нека условно се "трансформираме" в център за данни: ще разгледаме няколко примера за това как се извършва параметризирането на задачите, събирането на изисквания и процеса на автоматично валидиране.



## ГЛАВА 4.4.

### СИСТЕМАТИЗИРАНЕ НА ИЗИСКВАНИЯТА И ВАЛИДИРАНЕ НА ИНФОРМАЦИЯТА

#### Събиране и анализ на изисквания: трансформиране на комуникациите в структурирани данни

Събирането и управлението на изискванията е първата стъпка към осигуряване на качеството на данните. Въпреки развитието на цифровите инструменти, повечето изисквания все още се формулират по неструктурен начин: чрез писма, протоколи от срещи, телефонни разговори и устни дискусии. Тази форма на комуникация затруднява автоматизирането, валидирането и повторното използване на информацията. В тази глава разглеждаме как да преведем текстовите изисквания във формални структури, като гарантираме, че бизнес изискванията са прозрачни и систематични.

Изследването на Gartner "Data Quality: Best Practices for Accurate Insights" (Качество на данните: Най-добри практики за точни прозрения) подчертава критичното значение на качеството на данните за успешни инициативи за данни и анализи [100]. Те отбелязват, че лошото качество на данните струва на организациите средно поне 12,9 млн. долара годишно и че надеждните, висококачествени данни са от съществено значение за създаването на компания, основана на данни.

Липсата на структурирани изисквания води до факта, че един и същ елемент (субект) и неговите параметри могат да се съхраняват в различни системи в различни варианти. Това не само намалява ефективността на процесите, но и води до загуба на време, дублиране на информация и необходимост от повторно потвърждаване на данните, преди да бъдат използвани. В резултат на това дори един пропуск - изгубен параметър или един неправилно описан елемент - може да забави вземането на решения и да доведе до неефективно използване на ресурсите.

Заради липсата на пирон подковата е изгубена.  
 Заради липсата на подкова конят е изгубен.  
 Заради липсата на кон ездачът е изгубен.  
 Заради липсата на ездач съобщението е изгубено.  
 Заради липсата на съобщение битката е загубена.  
 Заради липсата на битка е загубено царството.  
 И всичко това заради липсата на един пирон в подковата.

- Поговорка [101]

Анализът и събирането на изисквания за процеса на попълване и съхраняване на данни започва с идентифициране на всички заинтересовани страни. Точно както пословичната загуба на един пирон води до верига от критични последици, така и в бизнеса загубата на една заинтересована страна, едно пренебрежнато изискване или загубата дори на един параметър може да окаже значително въздействие не само върху отделен бизнес процес, но и върху цялата екосистема на проекта.

и организацията като цяло. Ето защо е изключително важно да се идентифицират дори онези елементи, параметри и роли, които на пръв поглед изглеждат незначителни, но впоследствие могат да се окажат критични за устойчивостта на бизнеса.

Нека си представим, че компанията има проект, по който клиентът поставя ново искане - "да се добави допълнителен прозорец от северната страна на сградата". Малкият процес "искане на клиента за добавяне на нов прозорец към настоящия проект" включва архитект, клиент, CAD специалист (BIM), строителен мениджър, мениджър логистика, ERP -аналитик, инженер по контрол на качеството, инженер по безопасност, мениджър контрол и мениджър собственост.

Дори малък процес може да включва десетки различни специалисти. Всеки участник в процеса трябва да разбира изискванията на специалистите, с които е свързан на ниво данни.

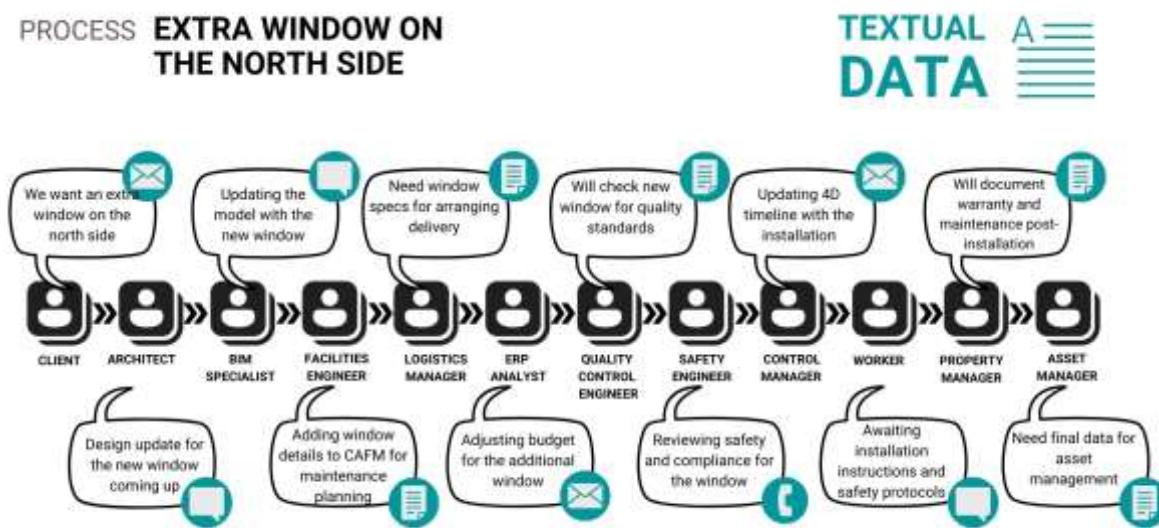
На ниво текст (фиг. 4.4-1) комуникацията между клиента и специалистите във веригата на процеса се осъществява по следния начин:

- ❶ **Клиент:** "Решихме да добавим допълнителен прозорец от северната страна за по-добро осветление. Може ли това да се осъществи?"
- ❷ **Архитект:** "Разбира се, ще преработя проекта, за да включа новия прозорец, и ще изпратя актуализирани CAD планове (BIM)".
- ❸ **Специалист по CAD (BIM):** "Получих нов проект. Актуализирам CAD (BIM) модела с допълнителния прозорец и след съгласуване с инженера по FEM предоставям точното местоположение и размери на новия прозорец".
- ❹ **Ръководител на строителството:** "Получен е нов проект. Коригираме датите за монтаж на 4D и информираме всички съответни подизпълнители".
- ❺ **Инженер по съоръженията (CAFM):** "Ще въведа данните от 6D на новия прозорец в системата CAFM за бъдещо управление на съоръженията и планиране на поддръжката."
- ❻ **Мениджър логистика:** "Трябват ми размерите и теглото на новия прозорец, за да организирам доставката на прозореца до обекта".
- ❼ **ERP -аналитик:** "Трябват ми таблиците с обхвата и точният тип прозорец за актуализация на бюджета 5D в нашата ERP система, за да отразя разходите за новия прозорец в общата оценка на проекта."
- ❽ **Инженер по контрол на качеството:** "След като спецификациите на прозорците са готови, ще се уверя, че те отговарят на нашите стандарти за качество и материали."
- ❾ **Инженер по безопасността:** "Ще оценявам аспектите на безопасността на новия прозорец, с особен акцент върху съответствието и евакуацията съгласно схемата 8D".
- ❿ **Ръководител на контрола:** "Въз основа на точния обхват на работата от ERP, ние ще актуализираме нашия график 4D, за да отразим инсталацирането на новия прозорец, и ще съхраним новите данни в системата за управление на съдържанието на проекта."
- ❰ **Работник (инсталатор):** "Необходими са ви инструкции за инсталациране, сглобяване и време за работа. Освен това има ли въведени специални правила за безопасност, които трябва да спазвам?"
- ❱ **Управител на имот:** "След като бъдат инсталирани, ще документирам информацията за гаранцията и поддръжката за дългосрочно управление".
- ❲ **Мениджър на активи:** "Инженер по оборудването, моля, изпратете окончателните данни за проследяване на активите и управление на жизнения им цикъл."

**Клиент:** "Чакайте, може би бързам и прозорецът няма да е необходим. Може би трябва да направя балкон".

При такива сценарии, които се случват често, дори малка промяна предизвиква верижна реакция между множество системи и роли. В този случай почти цялата комуникация в началната фаза на е в текстова форма: имейли, чатове, протоколи от срещи (фиг. 4.4-1).

В такава система за текстова комуникация за строителен проект е много важно да има система за правно потвърждение и записване на всички операции по обмен на данни и всички взети решения. Така се гарантира, че всяко взето решение, инструкция или промяна са правно валидни и проследими, което намалява риска от бъдещи "недоразумения"



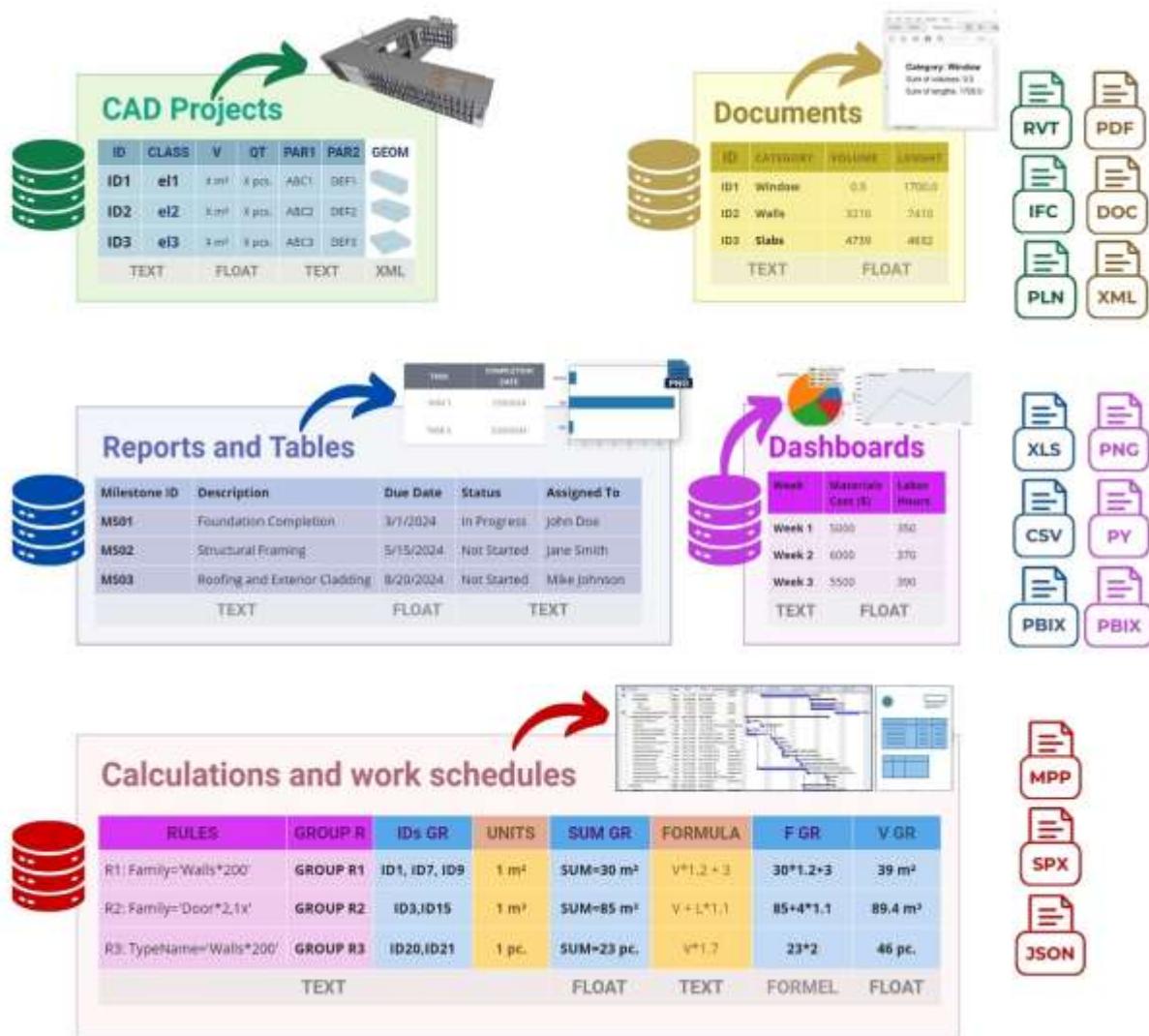
Фиг. 4.4-1 Комуникацията между клиента и изпълнителя в началните етапи на проекта често съдържа многоформатни текстови данни.

Липсата на правен контрол и валидиране на решенията в съответните системи на строителния проект може да доведе до сериозни проблеми за всички участници. Всяко решение, заповед или промяна, направени без подходящо документиране и валидиране, могат да доведат до спорове (и съдебни дела).

Правното консолидиране на всички решения в текстовата комуникация може да се осигури само чрез голям брой подписани документи, които ще паднат върху плещите на ръководството, което е задължено да регистрира всички сделки. В резултат на това, ако от всеки участник се изисква да подписва документи за всяко действие, системата губи гъвкавост и се превръща в бюрократичен лабиринт. Липсата на потвърждения на транзакциите не само ще забави изпълнението на проекта, но може да доведе и до финансови загуби и влошаване на отношенията между участниците, стигайки до правни проблеми.

Такъв процес на одобрение на транзакция, който обикновено започва с текстови дискусии, постепенно се превръща в обмен на многоформатни документи на следващите етапи (фиг. 4.4-2), което значително усложнява комуникацията, която досега се осъществяваше само чрез текст. Без ясно

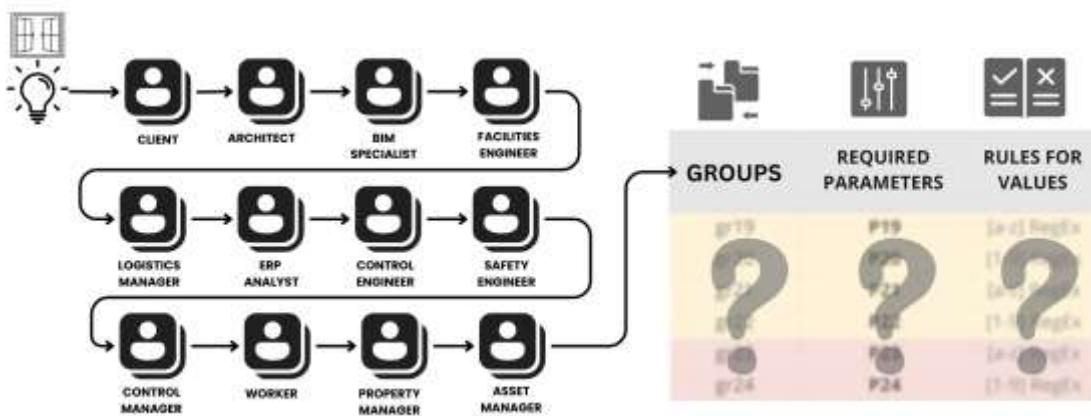
дифинирани изисквания автоматизирането на такива процеси, които са изпълнени с многоформатни данни и голям брой текстови изисквания, става почти невъзможно.



Фиг. 4.4-2 Всяка система в средата на строителната компания служи като източник на правно значими документи в различни формати.

Текстовите съобщения изискват всеки специалист да се запознае с пълната кореспонденция или да присъства редовно на всички срещи, за да разбере текущото състояние на проекта.

За да преодолее това ограничение, е необходимо да се премине от текстова комуникация към структуриран модел на изискванията. Това е възможно само чрез систематичен анализ, визуализация на процесите и описание на взаимодействията под формата на блок-схеми и модели на данни (фиг. 4.4-3). Също както при моделирането на данни (фиг. 4.3-7), от контекстуално-идейното ниво преминахме към концептуалното ниво, като добавихме системите и инструментите, използвани от участниците, и връзките между тях.



Фиг. 4.4-3 За да научите как да управлявате и автоматизирате процеса на валидиране, е необходимо да визуализирате процесите и да структурирате изискванията.

Първата стъпка в систематизирането на изискванията и връзките е да се визуализират всички връзки и взаимоотношения с помощта на концептуални блок-схеми. Концептуалното ниво не само ще улесни всички участници в процеса да разберат цялата верига на процеса, но и ясно ще покаже защо и за кого са необходими данните (и изискванията) на всеки етап от процеса.

## Блок-схеми на процеси и ефективност на концептуалните рамки

За да преодолеят пропастта между традиционните и модерните подходи към управлението на данни, компаниите трябва съзнателно да преминат от фрагментирани текстови описания към структурирани представяния на процесите. Еволюцията на данните - от глинени плочки до цифрови екосистеми - изисква нови инструменти за мислене. И един такъв инструмент е концептуалното моделиране с помощта на блок-схеми. Създаването на визуални диаграми - блок-схеми, диаграми на процесите, диаграми на взаимодействието - позволява на участниците в проекта да осъзнават как техните действия и решения влияят върху цялата система за вземане на решения.

Ако процесите изискват не само съхраняване на данни, но и тяхното анализиране или автоматизиране, тогава трябва да започнете да се занимавате с темата за създаване на концептуален и визуален слой с изисквания.

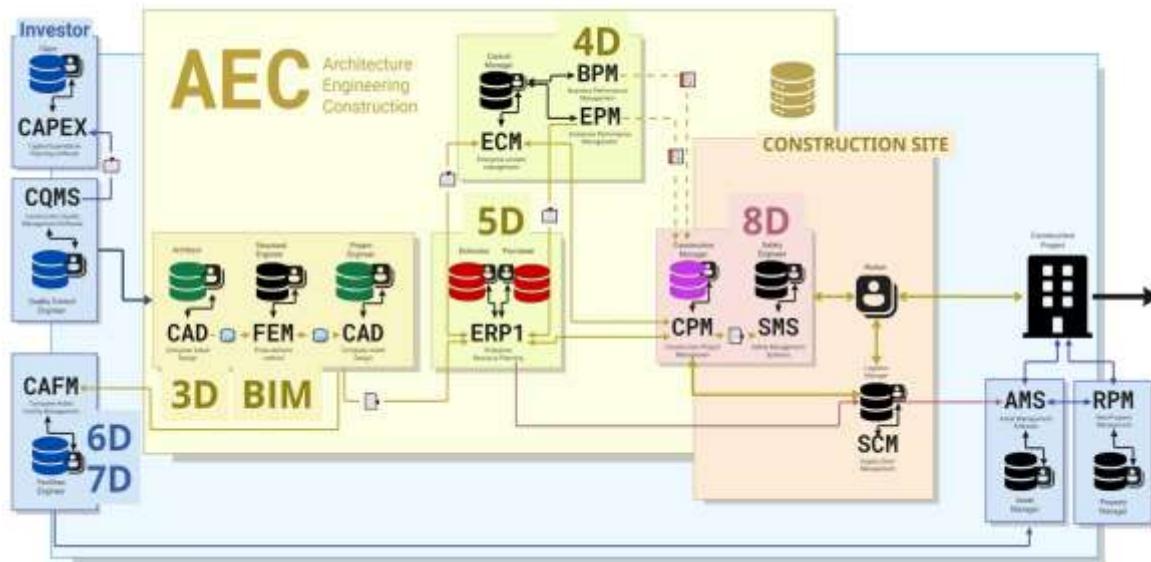
В нашия пример (фиг. 4.4-1) всеки специалист може да бъде част от малък екип, но също така и от по-голям отдел, включващ до десетина експерти под контрола на главен мениджър. Всеки отдел използва база данни на специализирано приложение (фиг. 1.2-4, например ERP, CAD, MEP, CDE, ECM, CRM и т.н.), която редовно се актуализира с постъпващата информация, необходима за създаване на документи, регистриране на правния статус на решенията и управление на процесите.

Процесът на транзакция е подобен на работата на древните мениджъри преди 4000 години, когато за юридическо потвърждаване на решенията са се използвали глинени плочки и папироси. Разликата между съвременните системи и техните предшественици от глина и хартия е, че съвременните методи допълнително включват процеса на преобразуване на текстовата информация в цифрова форма за по-нататъшна автоматична обработка в други системи и инструменти.

Създаването на визуализация на процеса под формата на концептуални блок-схеми ще ви помогне да опишете всяка стъпка и взаимодействията между различните роли, правейки сложния работен процес ясен и прост.

Визуализирането на процесите в гарантира, че логиката на процеса е прозрачна и достъпна за всички членове на екипа.

Същият комуникационен процес за добавяне на прозорец към проекта, който беше описан под формата на текст, съобщения (Фигура 4.4-1) и блокова диаграма, е подобен на концептуалния модел, който разглеждахме в главата за моделиране на данни (Фигура 4.4-4).



Фиг. 4.4-4 Концептуалната диаграма показва участниците в проекта като потребители на база данни, в която заявките им свързват различни системи.

Въпреки че концептуалните диаграми са важна стъпка, много компании се ограничават до това ниво, като смятат, че визуалната диаграма е достатъчна за разбиране на процесите. Това създава илюзията за управляемост: мениджърите могат по-лесно да възприемат цялостната картина на такава блок-схема и да видят връзките между участниците и етапите. Такива схеми обаче не дават ясна представа за това какви данни са необходими за всеки участник, в какъв формат трябва да се предават и какви параметри и атрибути са задължителни за реализирането на автоматизацията. Концептуалната блок-схема прилича по-скоро на маршрутна карта: тя показва кой с кого взаимодейства, но не разкрива какво се предава при тези взаимодействия.

Дори ако даден процес е описан подробно на концептуално ниво с помощта на блок-схеми, това не гарантира неговата ефективност. Визуализацията често опростява работата на мениджърите, като им позволява по-лесно да проследяват процеса с доклади стъпка по стъпка. За инженерите по бази данни обаче концептуалното представяне може да не е достатъчно ясно и да не дава ясна представа за това как да се реализира процесът на ниво параметри и изисквания.

С преминаването към по-сложни екосистеми от данни първоначалното внедряване на концептуални и визуални инструменти става от решаващо значение, за да се гарантира, че процесите, свързани с данните, са не само ефективни, но и съгласувани със стратегическите цели на организацията. За да преведем напълно този процес на добавяне на прозорци (фигура 4.4-1) на ниво изисквания към данните, трябва да отидем едно ниво по-дълбоко и да преведем концептуалната визуализация на процеса на логическо и физическо ниво на данните, изискваните атрибути и техните гранични стойности.

## Структурирани изисквания и регулярни изрази RegEx

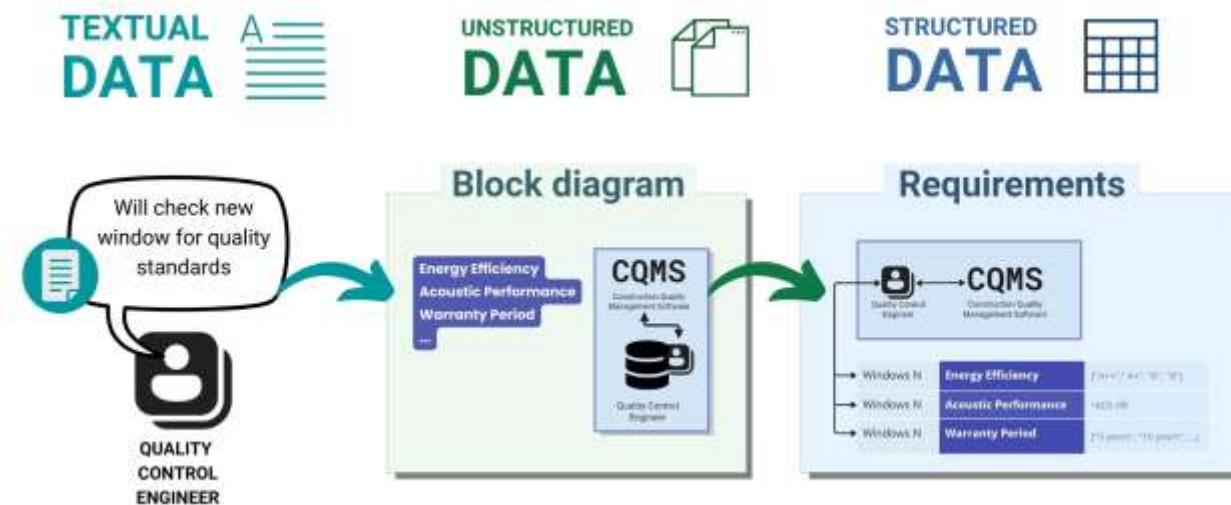
До 80% от данните, създадени в компаниите, са в неструктурирани или полуструктурирани формати [52] - текст, документи, писма, PDF -файлове, разговори. Такива данни (фиг. 4.4-1) е трудно да се анализират, проверяват, прехвърлят между системите и използват в автоматизацията.

За да се осигури управляемост, прозрачност и автоматично валидиране, е необходимо текстовите и полуструктуриранны изисквания да се трансформират в добре дефинирани, структурирани формати. Процесът на структуриране се отнася не само до данните (които разглеждахме подробно в първите глави на тази част от книгата), но и до самите изисквания, които участниците в проекта обикновено формулират в свободен текстов вид през целия жизнен цикъл на проекта, често без да мислят, че тези процеси могат да бъдат автоматизирани.

Точно както вече преобразувахме данните от неструктуррирана текстова форма в структурирана форма, в работния процес на изискванията ще преобразуваме текстовите изисквания в структуриран формат на "логически и физически слой".

Като част от примера за добавяне на прозорец (Фигура 4.4-1), следващата стъпка е да се описват изискванията за данните в табличен вид. Ще структурираме информацията за всяка система, използвана от участниците в проекта, като посочим ключовите атрибути и техните гранични стойности

Да разгледаме например една такава система (фиг. 4.4-5) - Системата за управление на качеството в строителството (CQMS), която се използва от инженера по контрол на качеството от страна на клиента. С нейна помощ той проверява дали даден нов елемент от проекта - в случая "нов прозорец" - отговаря на установените стандарти и изисквания.



Фиг. 4.4-5 Конвертирането на текстови изисквания в табличен формат с описания на атриутите на същността опростява разбирането им от други специалисти.

Като пример може да се разгледат някои важни изисквания за атриутите на обекти от тип "прозоречни системи" в CQMS -система (фиг. 4.4-6): енергийна ефективност, акустични характеристики и гаранционен срок. Всяка категория включва определени стандарти и спецификации, които трябва да се вземат предвид при проектирането и монтажа на прозоречни системи.



Фиг. 4.4-6 Инженерът по контрол на качеството трябва да провери новите елементи от типа прозорец за енергийна ефективност, звукоизолация и гаранционни стандарти.

Изискванията за данни, които инженерът по контрол на качеството определя под формата на таблица, имат например следните гранични стойности:

- **Класът на енергийна ефективност на прозорците** варира от "A++", който означава най-висока ефективност, до "B", считан за минимално приемливо ниво, като тези класове са представени със списък от приемливи стойности ["A++","A+","A","A","A","B"].
- **Акустичната изолация на прозорците**, измерена в децибели и показваща способността им

да намаляват уличния шум, се определя с правилния израз  $\backslash d\{2\}dB$ .

- **Атрибутът "Гаранционен срок"** за същността "Вид прозорец" започва от пет години, като определя този срок като минимално допустим при избора на продукт; задават се и стойности на гаранционния срок като **"5 години", "10 години" и т.н.** или логическото условие **">5 (години)"**.

Съгласно събранныте изисквания, в рамките на установените атрибути, нови елементи от категория или клас прозорци с оценки под "B", като "C" или "D", няма да преминат успешно теста за енергийна ефективност. Акустичната изолация на прозорците в данните или документите до инженера по качеството трябва да бъде обозначена с двуцифreno число, последвано от постфикс "dB", като например "35 dB" или "40 dB", а стойности извън този формат, като например "9 D B" или "100 децибела", няма да бъдат приети (тъй като няма да преминат модела за RegEx низове). Гаранционният срок трябва да започва с минимум "5 години", а прозорци с по-кратки гаранционни срокове, като например "3 години" или "4 години", няма да отговарят на изискванията, които инженерът по качеството е описал във формата на таблицата.

За да проверим такива стойности на атрибут-параметър спрямо граничните стойности от изискванията в процеса на валидиране, използваме или списък с позволени стойности (`["A", "B", "C"]`), или речници (`["A": "H1", "H2"; "B": W1, "W2"]`), логически операции (напр. `>`, `<`, `<=`, `>=`, `==`) за числови стойности) и регулярни изрази (за символни низове и текстови стойности, като например в атрибута `"Acoustic Performance"`). Регулярните изрази са изключително важен инструмент при работа със символни стойности.

**Регулярните изрази** (RegEx) се използват в езиците за програмиране, включително Python (библиотека `re`), за намиране и модифициране на низове. Regex е като детектив в света на символните низове, който може да идентифицира с точност текстови модели в текста.

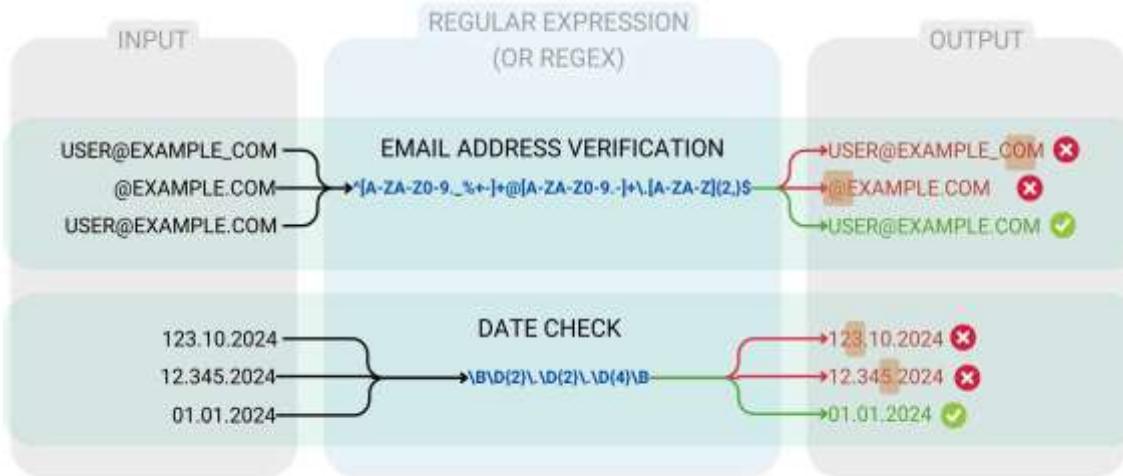
В регулярните изрази буквите се описват директно чрез съответните символи от азбуката, докато числата могат да се представят чрез специалния символ `\d`, който съответства на всяка цифра от 0 до 9. Квадратни скоби се използват за обозначаване на диапазон от букви или цифри, например `[a-z]` за всяка малка буква от латинската азбука или `[0-9]`, което е еквивалентно на `\d`. За нецифрови и небуквени символи се използват съответно `\D` и `\W`.

Популярни случаи на използване на RegEx (Фигура 4.4-7):

- **Проверка на имейл адрес**: за да проверите дали даден низ е валиден имейл адрес, можете да използвате шаблона `^\w+@\w+\.\w{2,}$`.
- **Извличане на дата**: Шаблонът `\b\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{2}\d{4}\b` може да се използва за извлечане на дата от текст във формат DD.MM.YYYY.
- Проверка на телефонни номера: за проверка на телефонни номера във формат +49(000)000-0000 моделът ще изглежда като `\+\d{2}(\d{3})\d{3}-\d{4}`.

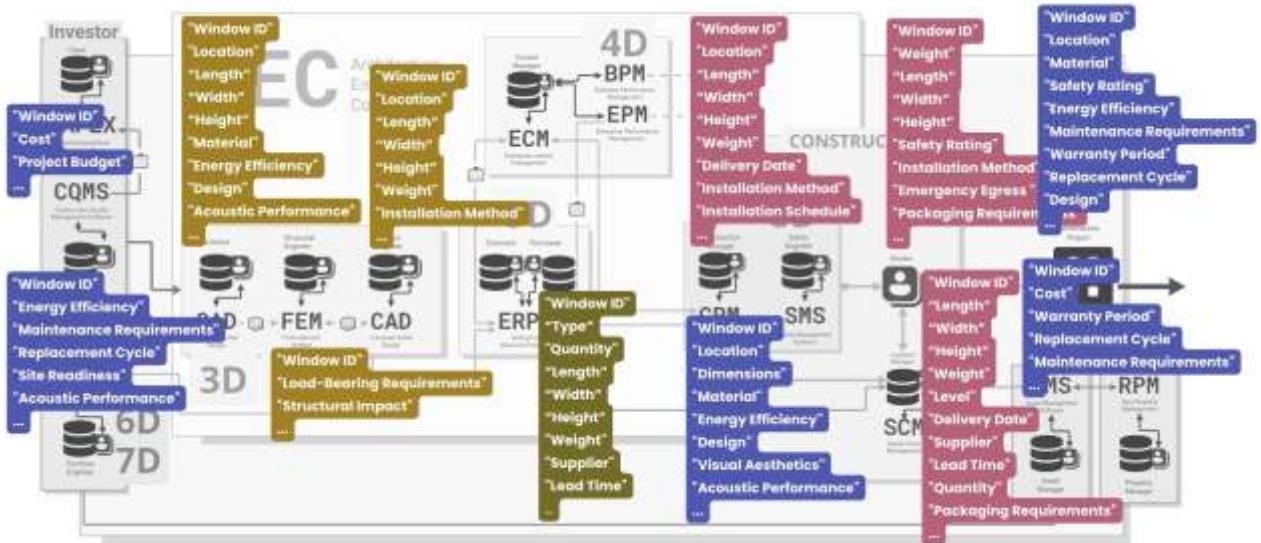
Като преведохме изискванията на инженер по качеството във формат на атрибути и техните гранични стойности (фиг. 4.4-6), ги превърнахме от първоначалния им текстов формат (разговори, писма и нормативни документи) в организирана и структурирана таблица, като по този начин да дохме възможност за автоматична проверка и анализ на всички постъпващи данни (напр. нови

елементи от категорията Window). Наличието на изисквания позволява данните, които не са били проверени, да бъдат автоматично отхвърлени, а проверените данни да бъдат автоматично прехвърлени към системите за по-нататъшна обработка.



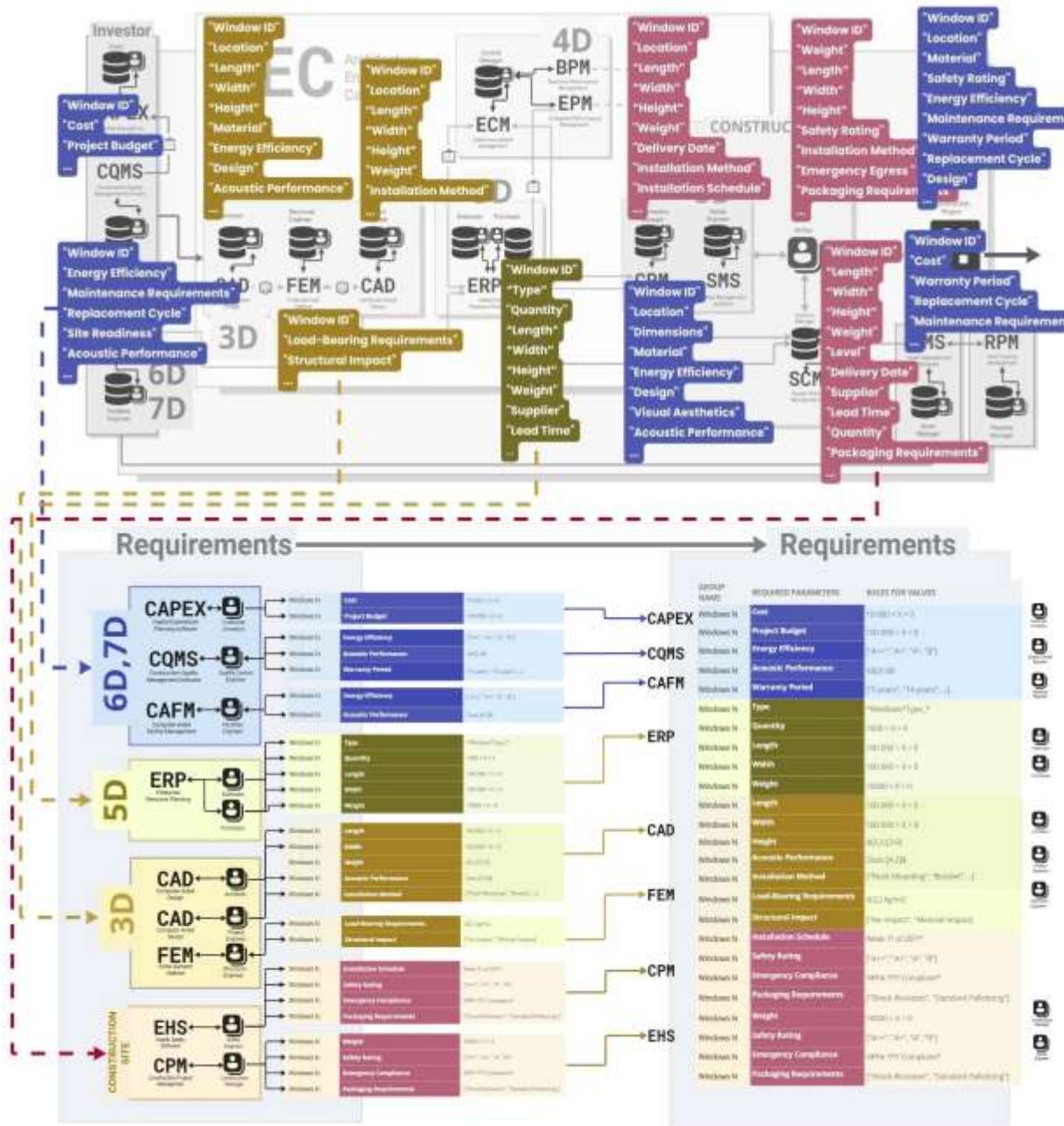
Фигура 4.4-7 Използването на регулярни изрази е изключително важен инструмент в процеса на валидиране на текстови данни.

Сега, преминавайки от концептуалното към логическото ниво на работа с изискванията, ще преобrazуваме всички изисквания на всички специалисти в нашия процес на инсталлиране на нов прозорец (фиг. 4.4-4) в подреден списък във формат на атрибути и ще добавим тези списъци с необходимите атрибути към нашата блок-схема за всеки специалист (фиг. 4.4-8).



Фигура 4.4-8 На ниво логически процес атрибутите, с които се занимава всеки специалист, се добавят към съответните им системи.

Чрез добавянето на всички атрибути към една обща таблица на процеса трансформираме информацията, представена преди това като текст и диалог на концептуално ниво (фиг. 4.4-1), в структурирана и систематизирана форма на таблици на физическо ниво (фиг. 4.4-9).



Фиг. 4.4-9 Конвертирането на неструктурнизирания диалог на специалистите в структурирани таблици помага да се разберат изискванията на физическо ниво.

Сега изискванията за данните трябва да бъдат съобщени на специалистите, които създават информация за конкретни системи. Например, ако работите в CAD база данни, преди да започнете да моделирате елементи, трябва да съберете всички необходими параметри въз основа на сценариите за крайно използване на данните. Обикновено това започва от оперативната фаза, следвана от строителната площадка, логистичния отдел, отдела за оценка, отдела за структурни изчисления и т.н. Едва след като сте взели предвид изискванията на всички тези области, можете да започнете да създавате данни - въз основа на събранныте параметри. Това ще ви позволи в бъдеще да автоматизирате проверката и прехвърлянето на данни по веригата.

Когато новите данни отговарят на изискванията, те автоматично се интегрират в екосистемата от данни на компанията и отиват директно при потребителите и системите, за които са предназначени. Проверката на данните спрямо атрибутите и техните стойности гарантира, че информацията отговаря на изискваните стандарти за качество и е готова за използване в сценариите на компанията.

Изискванията към данните са определени и сега, преди да започне проверката, данните, които трябва да бъдат проверени, трябва да бъдат създадени, получени или събрани, или текущото състояние на информацията в базите данни трябва да бъде записано, за да се използва в процеса на проверка.

## Събиране на данни за процеса на проверка

Преди да започне валидирането, е важно да се гарантира, че данните са налични във вид, подходящ за процеса на валидиране. Това не означава просто да разполагате с информацията, а да я подгответе: данните трябва да бъдат събрани и трансформирани от неструктурирани, слабо структурирани, текстови и геометрични формати в структурирана форма. Този процес е описан подробно в предишните глави, където са разгледани методите за трансформиране на различни видове данни. В резултат на всички трансформации входящите данни приемат формата на отворени структурирани таблици (фиг. 4.1-2, фиг. 4.1-9, фиг. 4.1-13).

Като имаме изискванията и структурирани таблици с необходимите параметри и гранични стойности (фиг. 4.4-9), можем да започнем валидирането на данните - или като единичен автоматизиран процес (Pipeline), или като поетапно валидиране на всеки входящ документ.

За да започне проверката, е необходимо или да се получи нов файл като вход, или да се фиксира текущото състояние на данните - да се създаде моментна снимка или да се експортират текущите и входящите данни, или да се установи връзка с външна или вътрешна база данни. В разглеждания пример такава моментна снимка се създава чрез автоматично преобразуване на CAD данни от структуриран формат, записан, да речем, в 23:00:00 часа в петък, 29 март 2024 г., след като всички проектанти са се прибрали у дома.



Фигура 4.4-10 Снимка на базата данни CAD (BIM), показваща текущата информация за атрибутите на нова единица от клас "Прозорец" в текущата версия на модела на проекта.

Благодарение на инструментите за обратен инженеринг, разгледани в главата "Превеждане на CAD данни (BIM) в структурирана форма", тази информация от различни CAD (BIM) инструменти и редактори може да се организира в отделни таблици (фиг. 4.4-11) или да се комбинира в една обща таблица, обединяваща различни части на проекта (фиг. 9.1-10).

Такава таблица - база данни показва уникални идентификатори на прозорци и врати (ID атриут), имена на типа (Type Name), размери (Width, Length), материали (Material), както и показатели за енергийни и акустични характеристики и други характеристики. Такава таблица, попълнена в CAD програмата (BIM), се събира от инженер-проектанта от различни отдели и документи, като формира информационен модел на проекта.

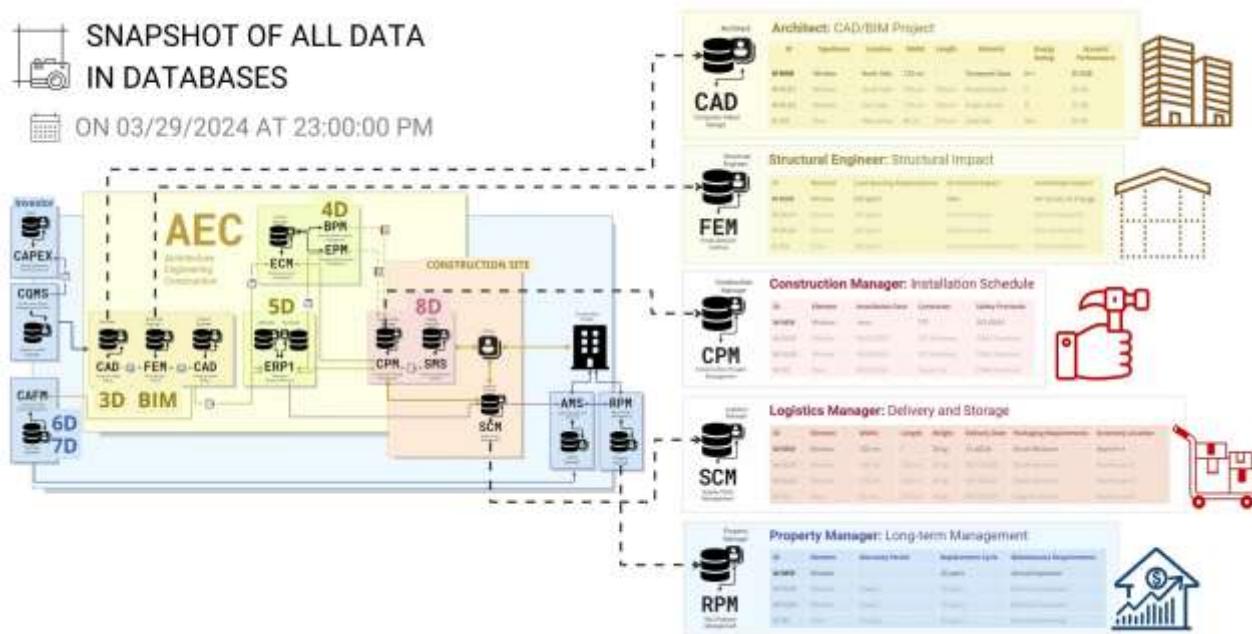


**Фигура 4.4-11 Структурираните данни от CAD системите могат да представляват двуизмерна таблица с колони, обозначаващи атрибутите на елементите.**

Истинските CAD (BIM) проекти включват десетки или стотици хиляди елементи (фиг. 9.1-10). Елементите в рамките на CAD форматите се категоризират автоматично по вид и категория - от прозорци и врати до плочи, плочи и стени. Уникалните идентификатори (напр. native ID, който се задава автоматично от CAD решението) или атрибутите на типа (Type Name, Type, Family) позволяват един и същ обект да бъде проследяван в различни системи. Например нов прозорец на северната стена на сграда може да бъде уникално идентифициран с един идентификатор "W-NEW" във всички съответни системи на организацията.

Въпреки че имената и идентификаторите на същностите трябва да бъдат последователни във всички системи, наборът от атрибути и стойности, свързани с тези същности, може да варира значително в зависимост от контекста на използване. Архитекти, строителни инженери, професионалисти в областта на строителството, логистиката и поддръжката на имоти възприемат едни и същи елементи по различен начин. Всеки от тях разчита на собствени класификатори, стандарти и цели: някои разглеждат прозореца единствено от естетическа гледна точка, оценявайки неговата форма и пропорции, докато други го разглеждат от инженерна или експлоатационна гледна точка, анализрайки топлопроводимостта, метода на монтаж, теглото или изискванията за поддръжка. Ето защо, когато се моделират данни и се описват елементи, е важно да се вземе предвид многостранното им използване и да се осигури съгласуваност на данните, като същевременно се отчитат спецификите на индустрията.

За всяка роля в процесите на компанията има специализирани бази данни със собствен потребителски интерфейс - от проектирането и изчисленията до логистиката, монтажа и експлоатацията на сградата (фиг. 4.4-12). Всяка такава система се управлява от професионален екип от специалисти чрез специален потребителски интерфейс или чрез заявки към базата данни, като зад сумата от всички решения, взети по въведените стойности в края на веригата, стои мениджърът на системата или мениджърът на отдела, който отговаря за юридическата валидност и качеството на въведените данни пред своите контрагенти, обслужващи други системи.



Фигура 4.4-12 Една и съща същност има един и същ идентификатор в различни системи, но различни атрибути, които са важни само в тази система.

След като сме организирали събирането на структурирани изисквания и данни на логическо и физическо ниво, остава да създадем процес за автоматична проверка на данните от различни входящи документи и различни системи спрямо предварително събранныте изисквания.

## Проверка на данните от и резултати от проверката

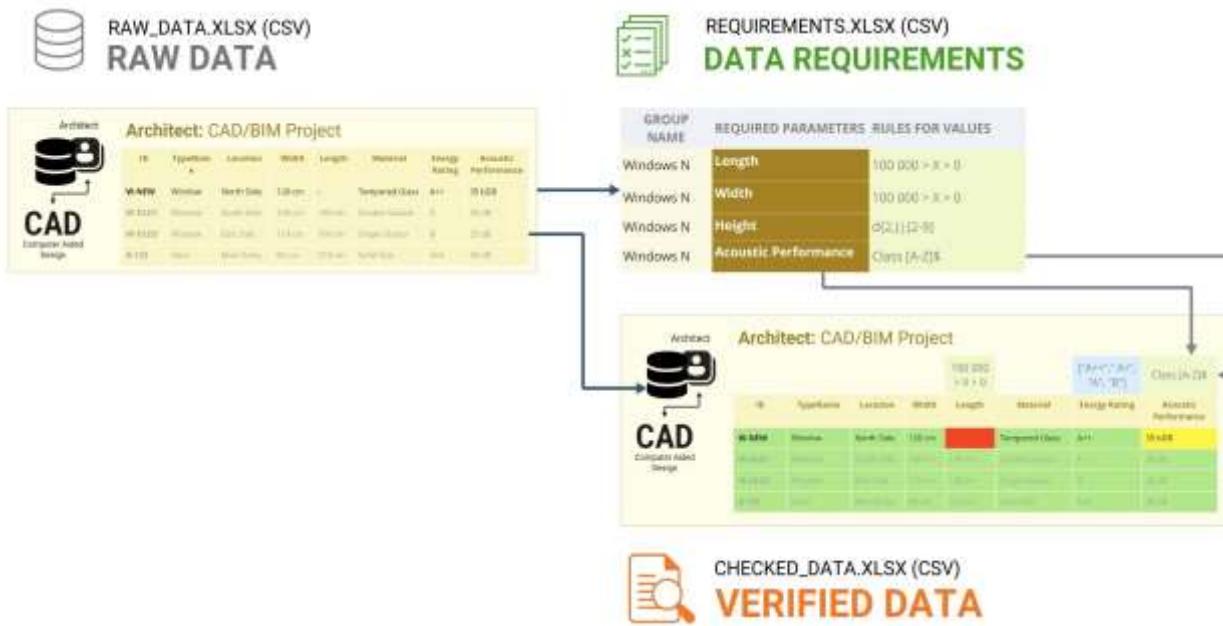
Всички нови данни, които постъпват в системата - независимо дали става въпрос за документи, таблици или записи в базата данни от клиента, архитекта, инженера, майстора, логистика или управителя на имота - трябва да бъдат валидирани спрямо формулираните по-рано изисквания (фиг. 4.4-9). Процесът на валидиране е от решаващо значение: всякакви грешки в данните могат да доведат до неправилни изчисления, забавяне на графика и дори до финансови загуби. За да се сведат до минимум тези рискове, трябва да се организира систематична и повтаряща се, итеративна процедура за валидиране на данните.

За да се валидират новите данни, които постъпват в системата - неструктурни, текстови или геометрични - те трябва да се преобразуват в слабо структуриран или структуриран формат. След това процесът на валидиране трябва да провери данните спрямо пълен списък от задължителни атрибути и техните допустими стойности.

Конвертирането на различни видове данни: текст, изображения, PDF -документи и смесени CAD (BIM) данни в структурирана форма е разгледано подробно в главата "Конвертиране на данни в структурирана форма".

Пример за това е таблица, получена от CAD (BIM) проект (фиг. 4.4-11). Тя включва полуструктурни геометрични данни и структурирана информация за атрибутите на проектните обекти (фиг. 3.1-14) - например елемент от клас "Windows".

За да извършим валидирането, сравняваме стойностите на атрибутите (фиг. 4.4-11) с референтните гранични стойности, които са определени от експертите под формата на изискване (фиг. 4.4-9). Окончателната сравнителна таблица (фиг. 4.4-13) ще покаже ясно кои стойности са приемливи и кои трябва да се коригират, преди данните да могат да се използват извън CAD приложенията (BIM).



Фигура 4.4-13 В окончателната таблица за валидиране се открояват онези стойности на атрибутите за новата същност от клас "Windows", на които трябва да се обърне внимание.

Прилагайки подобно решение с помощта на библиотеката Pandas, която описахме по-рано в главата "Pandas: Незаменим инструмент за анализ на данни", ще валидирате данните от табличен файл, извлечен от CAD файл (BIM) (RVT, IFC, DWG, NWS, DGN) (фигура 4.4-11), като използвате изискванията от друг табличен файл с изисквания (фигура 4.4-9).

За да получим кода, трябва да опишем в подкана за LLM, че трябва да заредим данните от файла

**raw\_data.xlsx** (пълен набор от данни от базата данни CAD (BIM)), да ги **проверим** и да запишем резултата в нов файл **checked\_data.xlsx** (фиг. 4.4-13).

- ❶ Нека получим кода, използвайки LLM, без да споменаваме библиотеката Pandas:

Напишете код, с който да проверите таблицата от файла raw\_data.xlsx и да ги валидирате, като използвате следните правила за валидиране: стойностите на колоните "Width" и "Length" са по-големи от нула, "Energy Rating" е включена в списъка ["A++", "A+", "A", "B"], а "Acoustic Performance" е променлива, която ще посочим по-късно - с добавена последна колона за валидиране, и запишете окончателната таблица в нов файл на Excel checked\_data.xlsx ↴

- ❷ В отговора на LLM ще бъде описан кратък пример за код на Python, който може да бъде усъвършенстван и допълнен чрез последващи подкани:

The screenshot shows a user interface for generating code. At the top right is a button labeled "Create in LLM Chat". Below it are four icons: a brain, a document, a star, and a whale. The main area contains a code editor window titled "Validation.py". The code is as follows:

```

1 import pandas as pd
2
3 # Load raw data from Excel file
4 df_data = pd.read_excel('raw_data.xlsx', sheet_name=0)
5
6 # Adding a 'Requirements Met' column that checks all conditions
7 df_data['Requirements Met'] = (
8     (df_data['Width'] > 0) &
9     (df_data['Length'] > 0) &
10    (df_data['Energy Rating'].isin(['A++', 'A+', 'A', 'B'])) &
11    (df_data['Acoustic Performance'] >= min_acoustic_performance)
12 )
13
14 # Save the results to a new Excel file
15 df_data.to_excel('checked_data.xlsx', index=False)

```

At the bottom right of the code editor is a "Run in IDE" button with icons for PyCharm, Visual Studio Code, and Jupyter Notebook.

Фигура 4.4-14 Кодът, генериран от LLM- модела, проверява конвертирация CAD (BIM ) проект спрямо изискванията за атрибути под формата на гранични стойности.

Кодът, генериран от езиковия модел LLM, може да се използва във всяка популярна IDE или онлайн инструмент: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с PyDev плъгин, Thonny, Wing IDE, IntelliJ IDEA с Python плъгин, JupyterLab или популярни онлайн инструменти Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Изпълнението на кода (фиг. 4.4-14) ще покаже, че "елементите на същността" W-OLD1, W-OLD2, D-

122 (и други елементи) от базата данни CAD (BIM) отговарят на изискванията за атрибутите: ширината и дължината са по-големи от нула, а класът на енергийна ефективност е една от стойностите в списъка "A++", "A", "B", "C" (фиг. 4.4-15).

Елементът W-NEW, от който се нуждаем и който насконо беше добавен и който отговаря за новия клас елементи "Window" от северната страна, не е съвместим (атрибут "Requirements Met"), тъй като дължината му е нула (стойност "0,0" се счита за неприемлива съгласно правилото ни за "Width">0) и не посочва клас на енергийна ефективност.



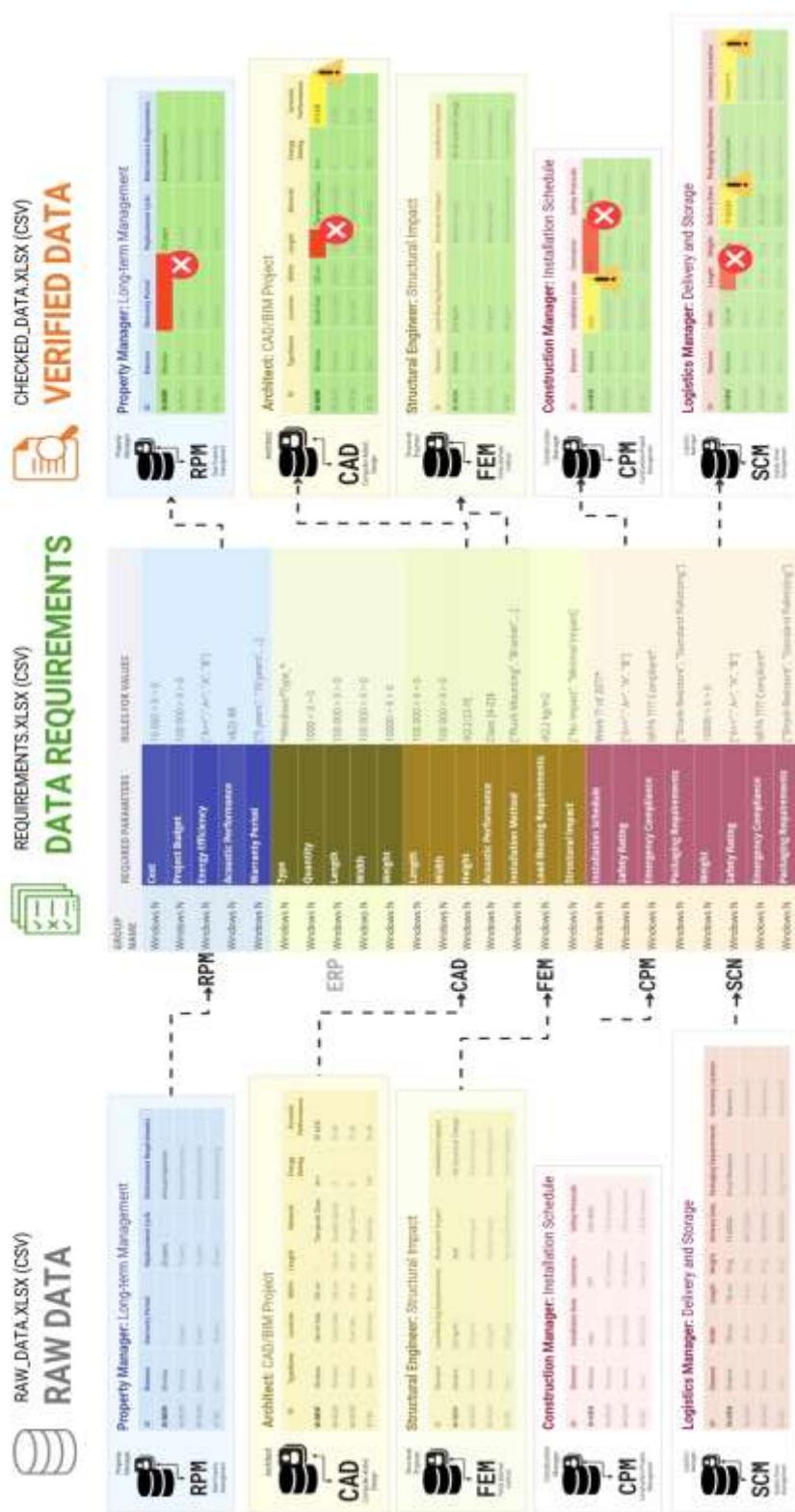
	ID	TypeName	Location	Width	Length	Material	Energy Rating	Acoustic Performance	Requirements Met
0	W-NEW	Window	North Side	120	0.0	Tempered Glass			False
1	W-OLD1	Window	South Side	100	140.0	Double Glazed	A++		True
2	W-OLD2	Window	East Side	110	160.0	Single Glazed	B		True
3	D-122	Door	Main Entry	90	210.0	Solid Oak	B-		True

**Фигура 4.4-15 Проверката идентифицира същности, които не са преминали през процеса на проверка, и добавя нов атрибут със стойности "False" (Невярно) или "True" (Вярно) към резултатите.**

По подобен начин проверяваме съответствието на всички елементи на проекта (същности) и необходимите атрибути за всяка от системите, таблиците или базите данни във всички данни, които получаваме от различни специалисти (фигура 4.4-1) по време на процеса на добавяне на прозорец към проекта.

В крайната таблица е удобно да се подчертаят резултатите от проверката в цвят за визуализация: в зелено се отбелязват атрибутите, които са преминали успешно проверката, в жълто - стойностите с некритични отклонения, а в червено - критичните несъответствия (фиг. 4.4-16).

В резултат на валидирането (Фигура 4.4-16) получаваме списък с надеждни и валидирани елементи с техните идентификатори, за които е проверено, че отговарят на изискванията за атрибути. Валидиралите елементи дават увереност, че тези елементи отговарят на обявените стандарти и спецификации за всички системи, участващи в процеса на добавяне на елементи от класа Window (Прозорец) или друг клас (ще обсъдим повече за автоматизирането на валидирането на данни и създаването на автоматизиран процес на ETL в главата "Автоматизиране на ETL и валидиране на данни").



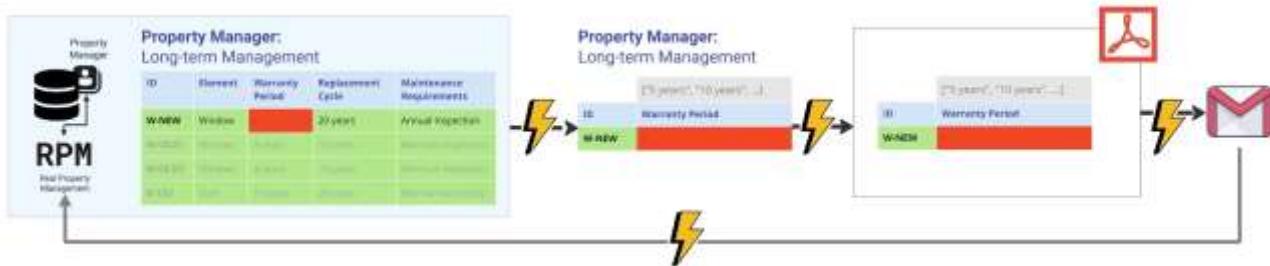
Фиг. 4.4-16 Резултатът от проверката, извършена за всички системи, позволява да се определи кои данни не отговарят на изискванията на компанията.

Субектите, които са били успешно валидирани, обикновено не се нуждаят от особено внимание. Те преминават към следващите етапи на обработка и интегриране в други системи без пречки. За разлика от "качествените" елементи, най-голям интерес представляват елементите, които не са успели да се валидират. Информацията за такива отклонения е от решаващо значение: тя трябва да се предава не само в таблични отчети, но и с помощта на различни инструменти за визуализация. Графичното представяне на резултатите от проверката помага за бърза оценка на цялостното състояние на качеството на данните, идентифициране на проблемните области и своевременно предприемане на коригиращи или коригирани действия.

## Визуализация на резултатите от проверката на

Визуализация е основен инструмент за интерпретиране на резултатите от проверката. В допълнение към обичайните обобщаващи таблици, той може да включва информационни табла, диаграми и автоматично генерираани PDF документи, които групират елементите на проекта по техния статус на инспекция. Цветното кодиране може да играе поддържаща роля тук: зеленото може да показва елементи, които са били успешно валидирани, жълтото може да показва елементи, които изискват допълнително внимание, а червеното може да показва елементи, които имат критични грешки или липсват ключови данни.

В нашия пример (фиг. 4.4-1) анализираме данните от всяка система стъпка по стъпка: от CAD (BIM) и управлението на собствеността до логистиката и графиците за монтаж (фиг. 4.4-16). След одита автоматично се генерираат индивидуални предупреждения или отчетни документи за всеки специалист, например в PDF формат (фиг. 4.4-17). Ако данните са коректни, специалистът получава кратко съобщение: "Благодарим ви за съвместната работа". Ако са открити несъответствия, се изпраща подробен доклад с текст: "В този документ са изброени елементи, техните идентификатори, атрибути и стойности, които не са проверени за съответствие".



**Фигура 4.4-17 Валидирането и автоматичното генериране на отчети ускоряват процеса на откриване и разбиране на недостатъците на данните за специалиста, който ги създава.**

Благодарение на автоматизирания процес на валидирание - веднага щом бъде открита грешка или пропуск в данните, на лицето, отговорно за създаването или обработката на съответните обекти и техните атрибути, се изпраща незабавно уведомление под формата на чат съобщение, електронна поща или PDF -документ (фиг. 4.4-18) със списък на елементите и описанията на атриутите, които не са били валидирани.



Фиг. 4.4-18 Автоматичните доклади от проверките улесняват разбирането на грешките и ускоряват работата по попълване на данните за проекта.

Например, ако в системата за управление на имоти (след структуриране) постъпи документ с неправилно попълнен атрибут "Гаранционен срок", управителят на имота получава предупреждение със списък на атрибутите, които трябва да бъдат проверени и коригирани.

По същия начин всякакви недостатъци в графика за инсталлиране или логистичните данни ще доведат до генериране на автоматичен доклад и например до изпращане на чат известие или имайл с резултатите от проверката до съответния специалист.

В допълнение към PDF -документи и графики с резултатите е възможно да се създават информационни табла и интерактивни 3D -модели (Фигура 7.1-6, Фигура 7.2-12), подчертаващи елементи с липсващи атрибути, което позволява на потребителите визуално да използват 3D геометриите на елементите, за да филтрират и оценяват качеството и пълнотата на тези елементи в проекта.

Визуализацията на резултатите от инспекциите на под формата на автоматично генеририани документи, графики или табла за управление значително опростява интерпретацията на данните и улеснява ефективната комуникация между участниците в проекта.

Процесът на автоматична проверка на данни от различни системи и източници на информация може да се сравни с вземането на информирани решения в ежедневието. Както компаниите в строителния бранш отчитат много променливи - от надеждността на входящите данни до тяхното въздействие върху сроковете, разходите и качеството на реализацията на проекта - така и при вземането на важни решения, например при избора на място за живееене, човек преценява цял набор от фактори: транспортна достъпност, инфраструктура, разходи, безопасност, качество на живот.

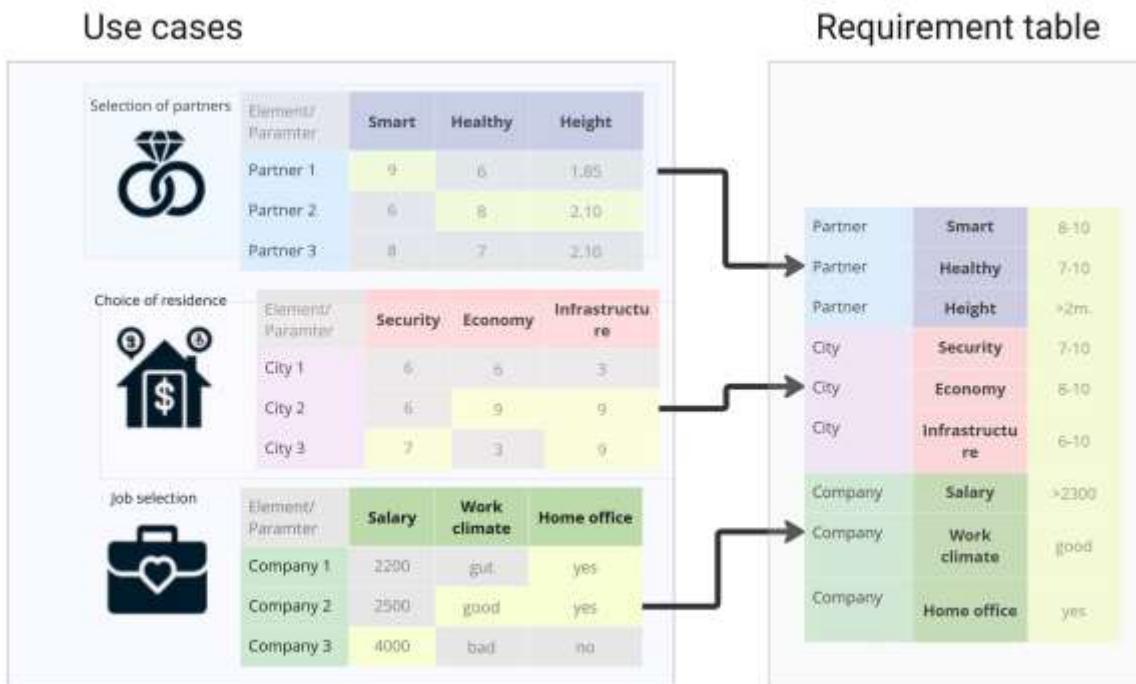
Всички тези съображения формират система от критерии, които са в основата на крайните решения, определящи нашия живот.

## Сравнение на проверките на качеството на данните с нуждите на човешкия живот

Въпреки постоянно развитието на методите и инструментите за контрол на качеството на данните, основният принцип на съответствие на информацията остава непроменен. Този принцип е вграден в основата на една зряла система за управление, независимо дали става въпрос за бизнес или за ежедневието.

Процесът на итеративно валидиране на данни много прилича на процеса на вземане на решения, с който всеки се сблъсква ежедневно. И в двата случая разчитаме на опит, данни и нова информация, когато тя стане достъпна. А все повече житейски и професионални решения - от стратегически до ежедневни - се вземат въз основа на данни.

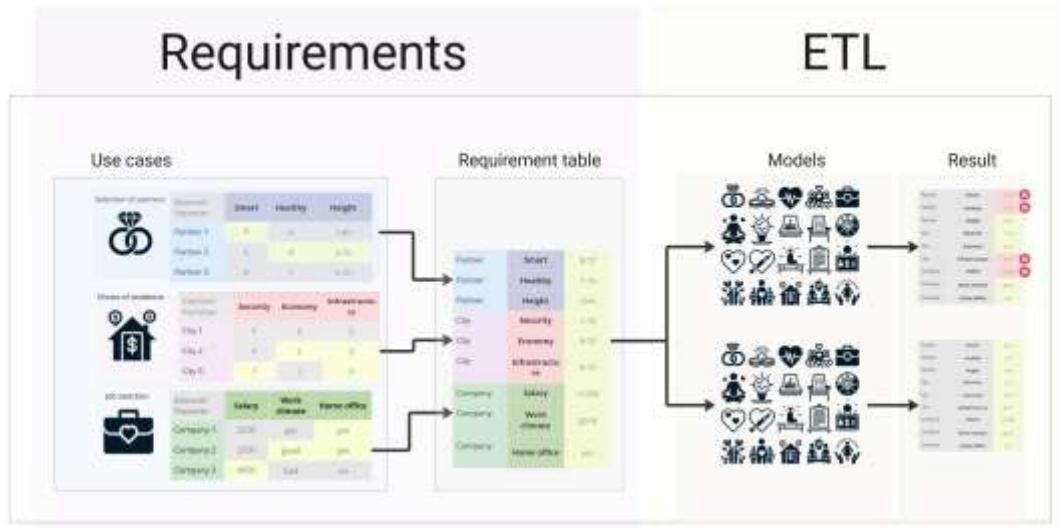
Например, когато избираме място за живееене или партньор в живота, ние интуитивно съставяме в съзнанието си таблица с критерии и характеристики, по които сравняваме алтернативите (фиг. 4.4-19). Тези характеристики - било то лични качества на човек или параметри на имот - представляват атрибути, които оказват влияние върху крайното решение.



Фиг. 4.4-19 Изборът на местоживееене, работа или партньорство се основава на изискванията за индивидуални атрибути.

Използването на структурирани данни и формализиран подход към описание на изискванията

(фигура 4.4-20) допринася за по-информиран и обоснован избор както в професионалния, така и в личния живот.

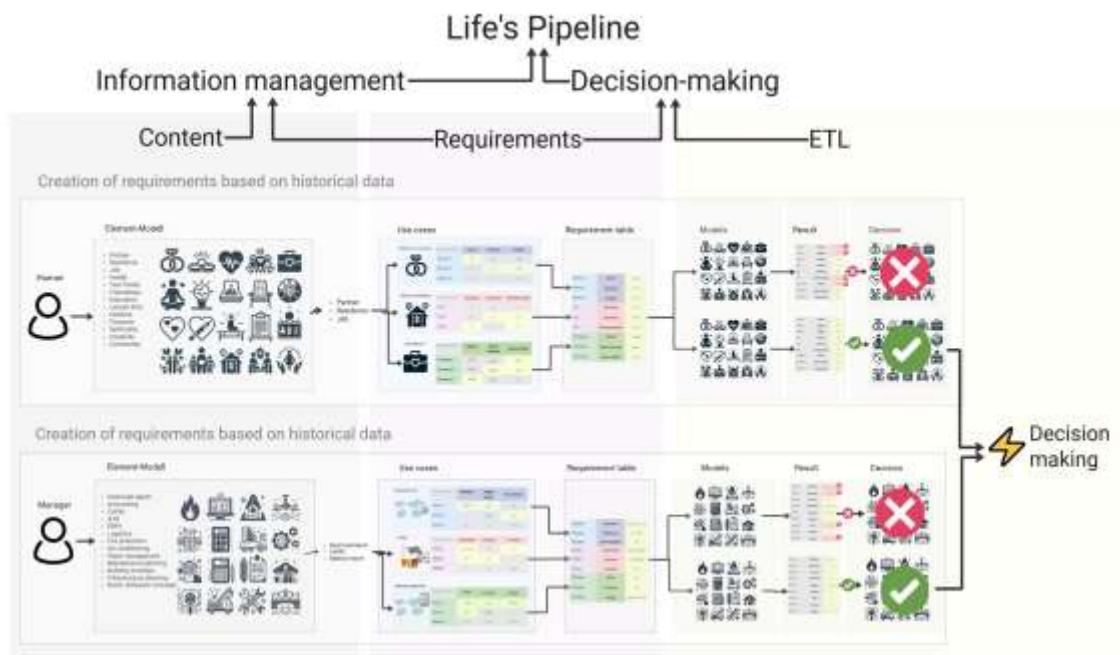


Фиг. 4.4-20 Формализирането на изискванията позволява да се систематизира възприемането на житейските и бизнес решения.

Подходът за вземане на решения, основан на данни, не е изключително бизнес инструмент. Той е безпроблемно интегриран и в ежедневието, като следва общи стъпки за обработка на данни (фиг. 4.4-21), подобни на процеса ETL (Extract, Transform, Load), който вече разглеждахме в началото на тази част при структурирането на данни и който ще разгледаме подробно в контекста на автоматизацията на задачи в седмата част на книгата:

- **Данните като основа (Извлечение):** Във всяка област - независимо дали става дума за работа или личен живот - ние събираме информация. В бизнеса това могат да бъдат до-клади, цифри, пазарни данни; в личния живот - личен опит, съвети от близки, отзиви, наблюдения.
- **Критерии за оценка (Трансформация):** събраната информация се интерпретира въз основа на предварително определени критерии. На работното място това са показатели за ефективност (KPI), бюджетни ограничения и норми; в личния живот - параметри като цена, удобство, надеждност, харизма и др.
- **Прогнозиране и анализ на риска (натоварване):** последният етап включва вземане на решение въз основа на анализ на трансформирани данни и сравняване на възможните последствия. Това е подобно на бизнес процесите, при които данните преминават през бизнес логика и филтър на риска.

Решенията, които вземаме - от тривиални предпочитания като това какво да ядем на закуска до важни житейски събития като избор на кариера или партньор в живота - по своята същност са резултат от обработка и оценка на данни.



**Фиг. 4.4-21 Бизнесът и животът като цяло са поредица от решения, основани на данни, при които качеството на данните, използвани за вземане на решения, е ключов фактор.**

Всичко в живота ни е взаимосвързано и както живите организми, включително хората, следват законите на природата, като се развиват и адаптират към променящите се условия, така и човешките процеси, включително начинът, по който събираме и анализираме данни, отразяват тези природни принципи. Тясната връзка между природата и човешката дейност потвърждава не само зависимостта ни от природата, но и желанието ни да прилагаме законите, усъвършенствани от милиони години еволюция, за да създаваме архитектури, процеси и системи за данни и вземане на решения.

Новите технологии, особено в строителството, са ярък пример за това как човечеството се вдъхновява от природата, за да създава по-добри, по-устойчиви и ефективни решения.

### Следващи стъпки: превръщане на данните в точни изчисления и планове

В тази част разглеждахме как да преобразуваме неструктурирани данни в структуриран формат, да разработваме модели на данни и да организираме процеси за проверка на качеството на информацията в строителните проекти. Управлението, стандартизацията и класификацията на данните е фундаментален процес, който изисква систематичен подход и ясно разбиране на бизнес изискванията. Техниките и инструментите, разгледани в тази част, позволяват надеждна интеграция между различни системи през целия жизнен цикъл на даден обект.

За да обобщим тази част, нека подчертаем основните практически стъпки, които ще ви помогнат да приложите обсъжданите подходи в ежедневните си задачи:

- Започнете със систематизиране на изискванията

Създаване на регистър с атрибути и параметри за ключови елементи на вашите проекти и процеси

Документиране на граничните стойности за всеки атрибут

Визуализиране на процесите и връзките между класове, системи и атрибути с помощта на блок-схеми (напр. в Miro, Canva, Visio)

#### ■ Автоматизиране на преобразуването на данни

Проверете кои от вашите документи, които често се използват в процесите, могат да бъдат дигитализирани с помощта на OCR библиотеки и да бъдат преобразувани в табличен вид.

Проверете инструментите за обратен инженеринг за извлечение на данни от CAD (BIM)

Опитайте се да настроите автоматично извлечение на данни от документи или формати, които често използвате в работата си, във вид на електронна таблица.

Настройка на автоматични преобразувания между различни формати на данни

#### ■ Създаване на база от знания за категоризиране

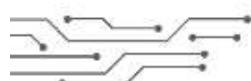
Разработване на вътрешен или използване на съществуващ класификатор на елементи, съобразен с индустрислните стандарти

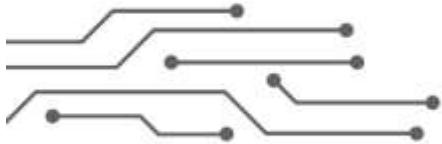
Документиране на взаимовръзките между различните системи за класификация

Обсъдете с екипа си темата за използването на единна система за идентификация и еднозначна класификация на елементите

Започнете да изграждате процес за автоматично валидиране на данните - както тези, с които работите в екипа, така и тези, които се предават на външни системи.

С помощта на тези подходи можете значително да подобрите качеството на данните си и да опростите последващата им обработка и преобразуване. В следващите части на тази книга ще разгледаме как да прилагаме вече структурирани и подгответни данни за автоматизирани изчисления, оценка на разходите, изготвяне на графики и управление на строителни проекти.





## V ЧАСТ

# ИЗЧИСЛЯВАНЕ НА РАЗХОДИТЕ И ВРЕМЕТО: ВКЛЮЧВАНЕ НА ДАННИ В СТРОИТЕЛНИТЕ ПРОЦЕСИ

Петата част е посветена на практическите аспекти на използването на данни за оптимизиране на разходите и планирането на строителни проекти. Подробно са анализирани ресурсно-базираният метод на оценяване на разходите и автоматизацията на процесите на оценяване. Разглеждат се методи за автоматизирано получаване на количествена сметка (от CAD (BIM)-модели и интегрирането им със системи за изчисляване. Разгледани са техниките за 4D и 5D моделиране за планиране на времето и управление на строителните разходи, с конкретни примери за тяхното приложение. Представен е анализ на разширени информационни слоеве 6D -8D, осигуряващи интегриран подход за оценка на устойчивостта, експлоатацията и безопасността на обекти на недвижимата собственост. Подробно са разгледани методите за изчисляване на въглеродния отпечатък и ESG -показателите на строителните проекти в контекста на съвременните екологични изисквания и стандарти. Критично се оценяват възможностите и ограниченията на традиционните ERP и PMIS системи в управлението на строителните процеси, като се анализира влиянието им върху прозрачността на ценообразуването. Прогнозирани са перспективите за преминаване от затворени решения към отворени стандарти и гъвкави инструменти за анализ на данни, способни да осигурят по-голяма ефективност на строителните процеси.

## ГЛАВА 5.1.

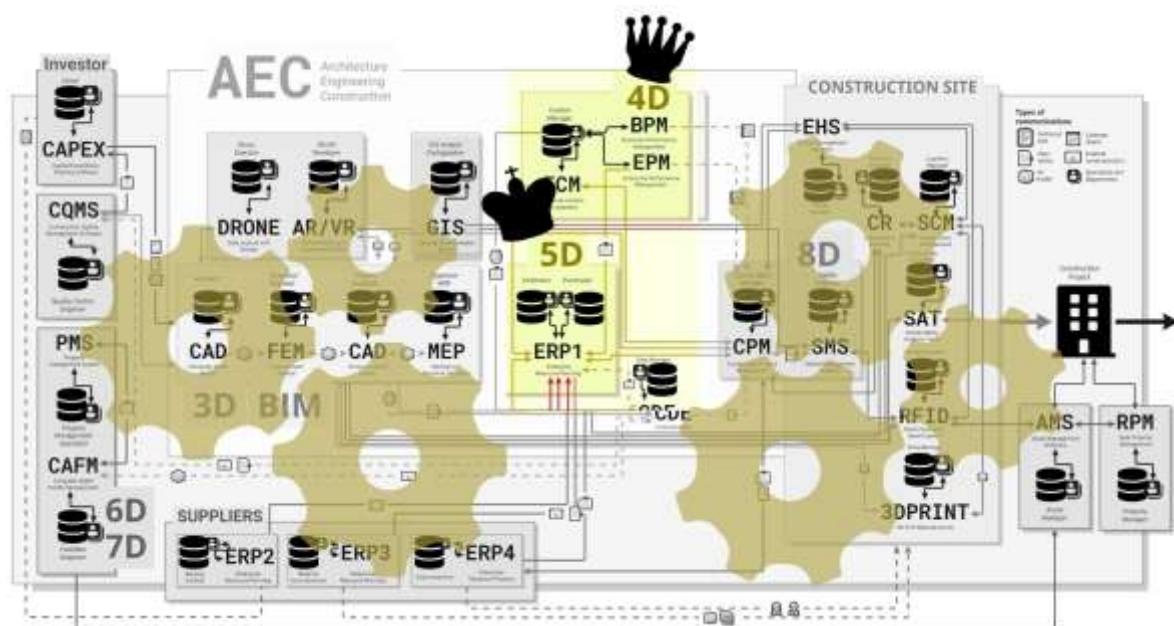
# ИЗЧИСЛЯВАНЕ НА РАЗХОДИ И ОЦЕНКИ ЗА СТРОИТЕЛНИ ПРОЕКТИ.

### Основи на строителството: оценяване на количеството, разходите и времето

Сред многото бизнес процеси, които определят устойчивостта на една компания в строителния бранш, от особена важност - както е било преди хиляди години - са процесите за точна оценка на броя на елементите, стойността на проекта и времето за изпълнение (фиг. 5.1-1).

Развитието на писмеността е резултат от редица фактори, сред които необходимостта от записване на икономическите транзакции, търговията и управлението на ресурсите в ранните общества. Първите правно значими документи - глинени плочки с изчисления на разходите за материали и труд - са били използвани в контекста на търговията и строителството. Тези плочки са записвали задълженията на страните при изграждането на структури и са били съхранявани като доказателство за споразумения и парични и търговски отношения.

В продължение на хилядолетия подходът към оценяването е останал до голяма степен непроменен: изчисленията са се извършвали ръчно, като се е разчитало на опита и интуицията на оценяващия инженер. С навлизането на модулните ERP системи и CAD инструментите обаче традиционният подход към оценяването на количествата, разходите и времето започна бързо да се трансформира. Съвременните цифрови технологии позволяват ключовите изчисления на времето и разходите да бъдат напълно автоматизирани, което позволява по-голяма точност, бързина и прозрачност при планирането на ресурсите на строителните проекти.



Фиг. 5.1-1 От многото различни системи най-важни в бизнеса са инструментите, които отговарят за показателите за обем, разходи и време.

Строителните компании се фокусират върху точните данни за времето и разходите. Те от своя страна зависят от количеството използвани материали и труд, а прозрачността им влияе върху рентабилността. Сложността на процесите на изчисление и липсата на прозрачност обаче често водят до по-високи разходи по проекта, неспазване на сроковете и дори до фалит.

Според доклада на KPMG "Познати проблеми - нови подходи" (2023 г.) само 50% от строителните проекти са завършени навреме, а 87% от компаниите съобщават за повишен контрол върху икономическите аспекти на капиталовите проекти. Основните проблеми са свързани с липсата на квалифициран персонал и трудностите при прогнозиране на риска [2].

Историческите данни за калкулирането на разходите и времето на процеса се събират по време на строителството на минали проекти през целия живот на строителната компания и се въвеждат в базите данни на различни системи (ERP, PMIS BPM, EPM и др.).

Наличието на качествени исторически данни за калкулиране на разходите е основно конкурентно предимство за строителната организация, което пряко влияе върху нейната жизнеспособност.

Отделите за оценяване и калкулиране на разходите в строителните и инженерните компании са създадени, за да събират, съхраняват и актуализират исторически данни за изчисленията на проекти. Тяхната основна функция е да натрупват и систематизират опита на компанията, което помага за подобряване на точността на оценката на обхвата, сроковете и разходите на нови проекти във времето. Този подход помага да се сведат до минимум грешките в бъдещите изчисления, базирани на практиката и резултатите от вече реализирани проекти.

## Методи за изчисляване на прогнозната стойност на проектите

Специалистите по изчисляване на разходите използват различни методи за оценка, като всеки от тях се фокусира върху определен вид данни, наличност на информация и ниво на детайлност на проекта. Най-често срещаните включват:

- **Метод, базиран на ресурсите:** оценяване на прогнозните разходи за даден проект въз основа на подробен анализ на всички необходими ресурси, като материали, оборудване и труд. Този метод изисква подробен списък на всички задачи и ресурсите, необходими за изпълнението на всяка задача, последван от изчисление на разходите за тях. Този метод е много точен и се използва широко в оценяването на разходите.
- **Параметричен метод:** използва статистически модели за оценка на разходите въз основа на параметрите на проекта. Това може да включва анализ на разходите за единица мярка, като например площ на сградата или обхват на работата, и адаптиране на тези стойности към конкретните условия на проекта. Методът е особено ефективен в ранните етапи, когато все още няма подробна информация.
- Метод на единица мярка (**метод на разходите за единица мярка**): изчислява прогнозните разходи за проекта въз основа на разходи за единица мярка (напр. за квадратен метър или кубичен метър). Това осигурява бърз и лесен начин за сравняване и анализиране на разходите за различни проекти или части от проекти.
- **Експертна преценка** (метод "Делфи"): основава се на мненията на експерти, които използват своя опит и знания, за да оценят стойността на даден проект. Подходът е полезен, когато не са налични точни изходни данни или проектът е уникален.

Струва си да се отбележи, че параметричният метод и експертните оценки могат да бъдат адаптирани към моделите за машинно обучение. Това дава възможност за автоматично генериране на прогнози за разходите и графика на проекта въз основа на обучителни извадки. Примери за прилагане на такива модели са разгледани по-подробно в главата Пример за използване на машинно обучение за намиране на разходите и графика на проекта" (фиг. 9.3-5).

Въпреки това методът, основан на ресурсите, остава най-популярният и широко използван в свидетелната практика. Той осигурява не само точна оценка на прогнозните разходи, но и позволява да се изчисли продължителността на отделните процеси на строителната площадка и на целия проект като цяло (повече подробности в главата "Строителни графики и 4D -проектни данни").

### Метод, базиран на ресурсите остойностяване и оценяване в строителството

Разходите, базирани на ресурсите, са метод на управленско счетоводство, при който разходите за даден проект се основават на прякото отчитане на всички участващи ресурси. В строителството този подход включва подробен анализ и оценка на всички материални, трудови и технически ресурси, необходими за извършване на работата.

Методът, основан на ресурсите, осигурява висока степен на прозрачност и точност при планирането на бюджета, тъй като се фокусира върху действителните цени на ресурсите към момента на оценката. Това е особено важно в нестабилна икономическа среда, където колебанията на цените могат да окажат значително влияние върху общите разходи по проекта.

В следващите глави ще разгледаме подробно процеса на ресурсно базирано калкулиране на разходите. За да разберем по-добре принципите му в строителството, ще направим аналогия с калкулирането на разходите за вечеря в ресторант. Мениджърът на ресторанта, планирайки вечерта, съставя списък на необходимите продукти, взема предвид времето за приготвяне на всяко ястие и след това умножава разходите по броя на гостите. В строителството процесът е подобен: за всяка категория елементи на проекта (обекти) се изготвят детайлни оценки Рецепти, а общата стойност на проекта се определя чрез сумиране на всички разходи в обща сметка - окончателната оценка по категории.

Ключовият и начален етап на подхода, основан на ресурсите, е създаването на първоначалната база данни на компанията. На първия етап от изчисляването на разходите се съставя структуриран списък на всички артикули, материали, видове работа и ресурси, с които компанията разполага в рамките на своите строителни проекти - от гвоздей в склада до описание на хората чрез тяхната квалификация и почасова ставка. Тази информация се систематизира в единна "База данни за строителни ресурси и материали" - табличен регистър, съдържащ данни за наименования, характеристики, мерни единици и текущи цени. Именно тази база данни се превръща в основен и първостепенен източник на информация за всички последващи изчисления на ресурсите - както на разходите, така и на сроковете за изпълнение на работите.

## **База данни за строителни ресурси: каталог на строителни материали и работи**

База данни или таблица на строителните ресурси и материали - включва подробна информация за всеки елемент, който може да се използва в строителен проект - продукт, артикул, материал или услуга, включително неговото наименование, описание, мерна единица и единична цена, записани в структуриран вид. В тази таблица можете да намерите всичко - от различните видове горива и материали, използвани в проектите, до подробни списъци на специалисти под формата на различни категории с описание на часовите ставки (фиг. 5.1-2).

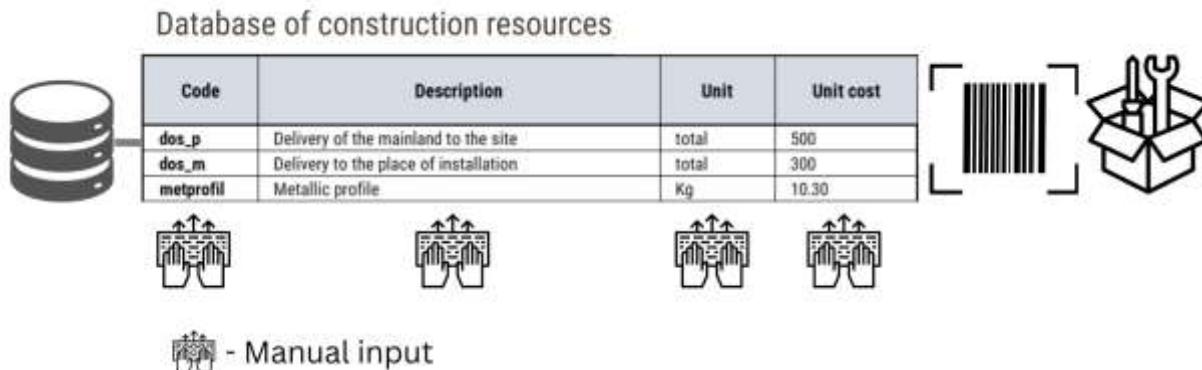
Database of resources	
 1st grade potatoes <b>1 kg   \$2,99</b>	 Sand lime bricks <b>1 pcs   \$1</b>
 Black Angus marble beef <b>1 kg   \$26,99</b>	 JCB 3CX backhoe loader <b>1 h   \$150</b>
 Broccoli <b>1 pcs   \$1,99</b>	 Laborer of the 1st category <b>1 h   \$30</b>

Фигура 5.1-2 Ресурсната таблица е списък със съставки, описващ материал и услуга с единична цена.

"Базата данни с ресурси" е подобна на продуктовия каталог на онлайн магазин, в който всеки артикул има подробно описание на своите атрибути. Това улеснява оценителите на разходите при избора на правилните ресурси (подобно на избора на продукти при добавянето им в пазарската кошница), необходими за изчисляване на конкретни строителни процеси под формата на калкулации (крайна поръчка в онлайн магазина).

Базата данни с ресурси може да се разглежда и като списък на всички съставки в готварска книга на ресторант. Всеки строителен материал, оборудване и услуга е подобен на съставките, използвани в рецептите. "Базата данни за ресурсите" е подробен списък на всички съставки - строителни материали и услуги, включително тяхната цена за единица: парче, метър, час, литър и т.н.

Нови елементи на същността могат да се добавят към таблицата "Бази данни за строителни ресурси" по два начина - ръчно (фиг. 5.1-3) или автоматично чрез интегриране със системите за управление на запасите на компанията или базите данни на доставчиците.



**Фиг. 5.1-3** Базата данни за ресурсите се попълва ръчно или автоматично се приемат данни от други бази данни.

Типична средна строителна компания използва база данни, съдържаща хиляди, а понякога и десетки хиляди артикули с подробни описания, които могат да се използват в строителните проекти. След това тези данни се използват автоматично в договорите и проектната документация, за да се опише точно съставът на работата и процесите

За да бъде в крак с променящите се пазарни условия, като например инфлацията, атрибутът "единична цена" за всеки продукт (стока или услуга) в базата данни за ресурсите (фигура 5.1-3) се актуализира редовно ръчно или чрез автоматично изтегляне на актуални цени от други системи или онлайн платформи.

Актуализирането на единичната цена на даден ресурс може да се извършва месечно, тримесечно или годишно - в зависимост от естеството на ресурса, инфлацията и външния икономически климат. Такива актуализации са необходими, за да се поддържа точността на изчисленията и оценките, тъй като това са основните елементи, които служат като отправна точка за работата на оценителите на разходите. Актуалните данни се използват за изготвяне на оценки, бюджети и графики, които отразяват реалните пазарни условия и намаляват риска от грешки в последващите изчисления по проекта.

## Изготвяне на изчисления и остойностяване на работите въз основа на ресурсна база

След като сте попълнили "Базата данни за строителни ресурси" (фиг. 5.1-3) с минимални единици, можете да започнете да създавате изчисления, които се изчисляват за всеки процес или работа на строителната площадка за определени мерни единици: например за един кубичен метър бетон, един квадратен метър стена от гипсокартон, за метър бордюр или за монтаж на прозорци.

Например, за изграждането на тухлена стена с площ 1 m<sup>2</sup> (Фигура 5.1-4), въз основа на опита от предишни проекти, са необходими приблизително 65 тухли (единица "Силикатна тухла") на цена 1 USD на бройка (атрибут "Цена на бройка"), общо 65 USD. Също така, според моя опит е необходимо да се използва строителна техника (entity "JCB 3CX Loader") в продължение на 10 минути, която ще постави тухли в близост до работната зона. Тъй като наемът на оборудването струва 150 USD на

час, 6 минути използване ще струват приблизително 15 USD. Освен това ще е необходим изпълнител за полагане на тухли за 2 часа, като часовата ставка е 30 USD, а общата сума е 60 USD.

	<b>x0.1</b> 1 kg   \$2,99		<b>x0.4</b> 1 kg.   \$2,99
	<b>x0.3</b> 1 kg   \$26,99		<b>x0.4</b> 1 kg.   \$26,99
	<b>x0.5</b> 1 pcs   \$1,99		<b>x65</b> 1 pcs   \$1
	+ margins and speculation		<b>x500</b> 1 pcs   \$1
	+ margins and speculation		<b>x0.1</b> 1 h   \$150
	+ margins and speculation		<b>x2</b> 1 h   \$150
	+ margins and speculation		<b>x2</b> 1 h   \$30
	+ margins and speculation		<b>x30</b> 1 h   \$30

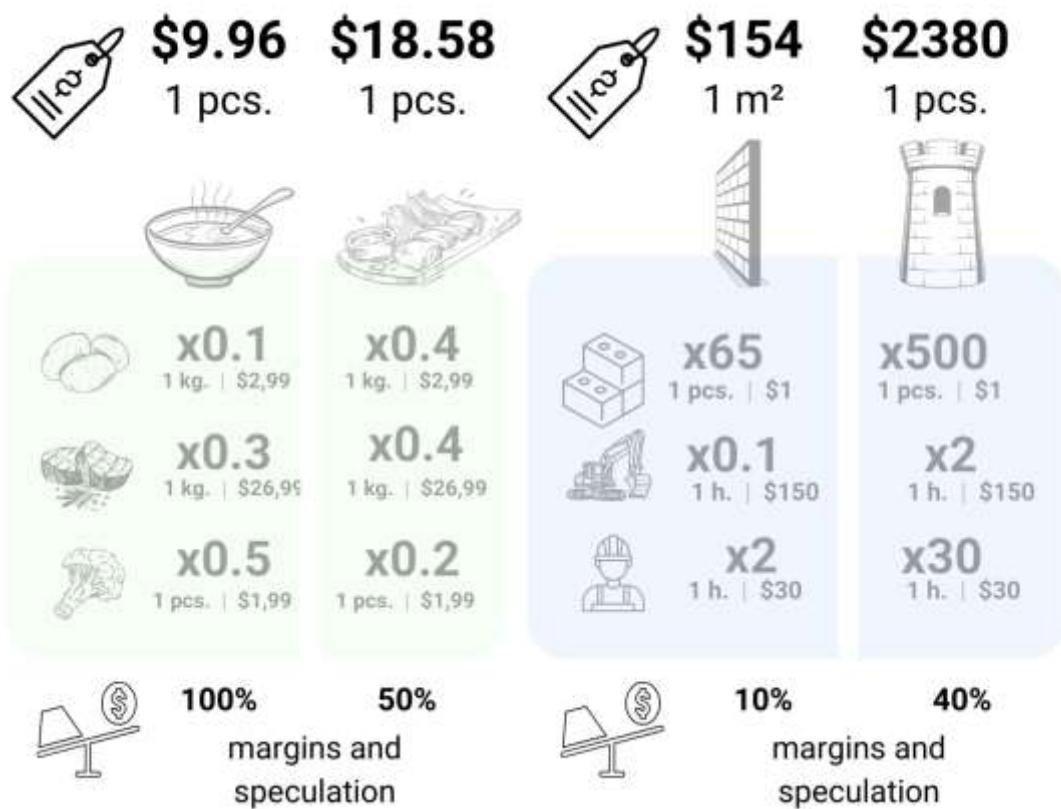
**Фиг. 5.1-4 Изчисленията на разходите предоставят подробен списък на строителните материали и услуги, необходими за изпълнение на строителните работи и процеси.**

Съставът на изчисленията (т.нар. "рецепти") се формира въз основа на историческия опит, натрупан от компанията в процеса на извършване на голям обем подобна работа. Този практически опит обикновено се натрупва чрез обратна връзка от строителната площадка. По-специално бригадирът събира информация директно на строителната площадка, като записва действителните разходи за труд, потреблението на материали и нюансите на технологичните операции. След това в сътрудничество с отдела за оценка тази информация се усъвършенства итеративно: прецизират се описанията на процесите, коригира се съставът на ресурсите и се актуализират калкулациите, за да се отразят действителните данни от последните проекти.

Точно както рецептата описва съставките и количествата, необходими за пригответо на дадено ястие, оценката на разходите съдържа подробен списък на всички строителни материали, ресурси и услуги, необходими за изпълнението на определена работа или процес.

Редовно извършваната работа позволява на работниците, бригадирите и оценителите да се ориентират в необходимото количество ресурси: материали, гориво, работно време и други параметри, необходими за извършване на единица работа (фиг. 5.1-5). Тези данни се въвеждат в системите за оценяване под формата на таблици, в които всяка задача и операция се описва чрез минималните

елементи на ресурсната база (с постоянно актуализирани цени), което гарантира точността на изчисленията.



Фиг. 5.1-5 За всяко задание се събират единични ставки, при които атрибутият за обем на единицата се умножава по нейното количество, като се добавя процент печалба.

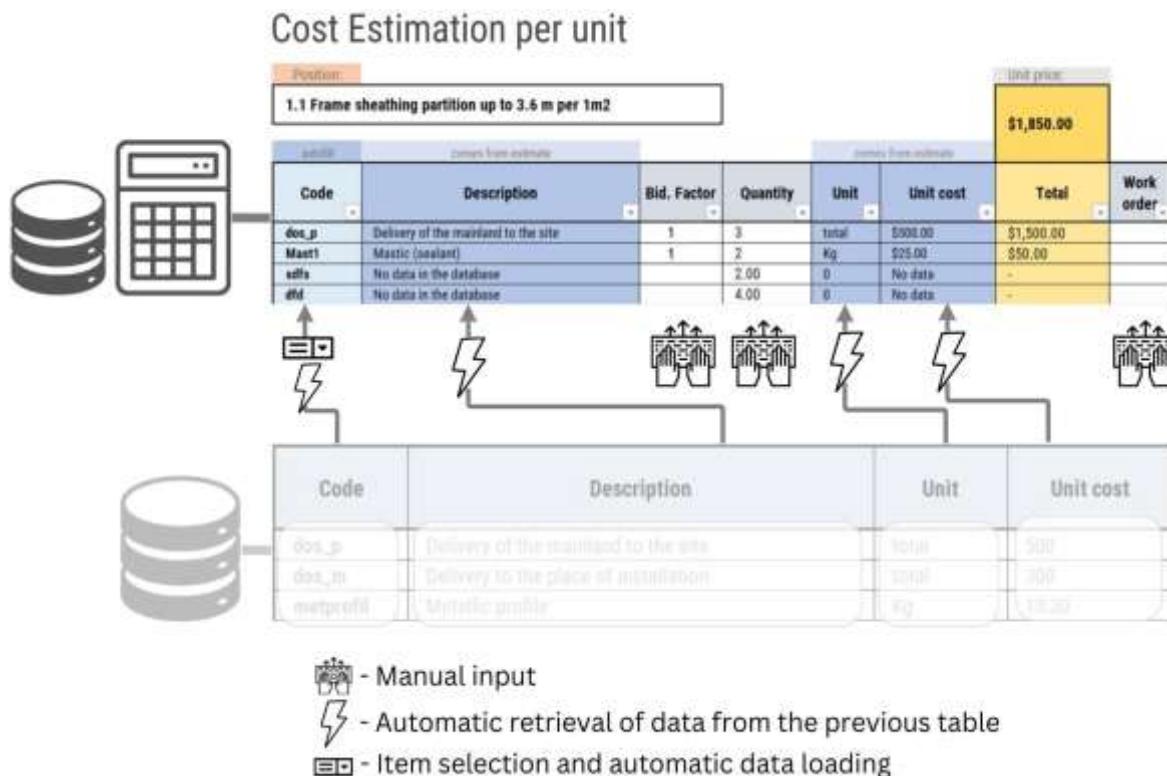
За да се получат общите разходи за всеки процес или дейност (обект на калкулиране), атрибутият на разходите се умножава по неговия брой и коефициенти. Коефициентите могат да отчитат различни фактори, като например сложността на работата, регионалните характеристики, процента на инфляция, потенциалните рискове (очакван процент на режийните разходи) или спекулациите (коефициент на допълнителна печалба).

Оценителят, в качеството си на анализатор, преобразува опита и препоръките на бригадира в стандартизиирани оценки, описващи строителните процеси чрез ресурсни единици в табличен вид. По същество задачата на оценителя е да събира и структурира чрез параметри и коефициенти информацията, идваща от строителната площадка.

По този начин крайните разходи за единица работа (напр. квадратен или кубичен метър, или един монтаж на единична единица) включват не само преките разходи за материали и труд, но и надценките на фирмата, режийните разходи, застраховките и други фактори (фигура 5.1-6)

В същото време вече не е необходимо да се притесняваме за реалните цени при изчисленията на

(рецептите), тъй като реалните цени винаги са отразени в "ресурсната база" (таблицата със съставките). На нивото на изчисленията данните от ресурсната база се зареждат автоматично в таблицата (напр. по кода на артикула или неговия уникален идентификатор), в която се зареждат описание и реалната цена за единица, която от своя страна може да се зареди автоматично от онлайн платформи или онлайн магазин за строителни материали. Оценителят на ниво изчисления работи трябва само да опише работата или процеса чрез атрибута "количество ресурси" и допълнителни фактори.



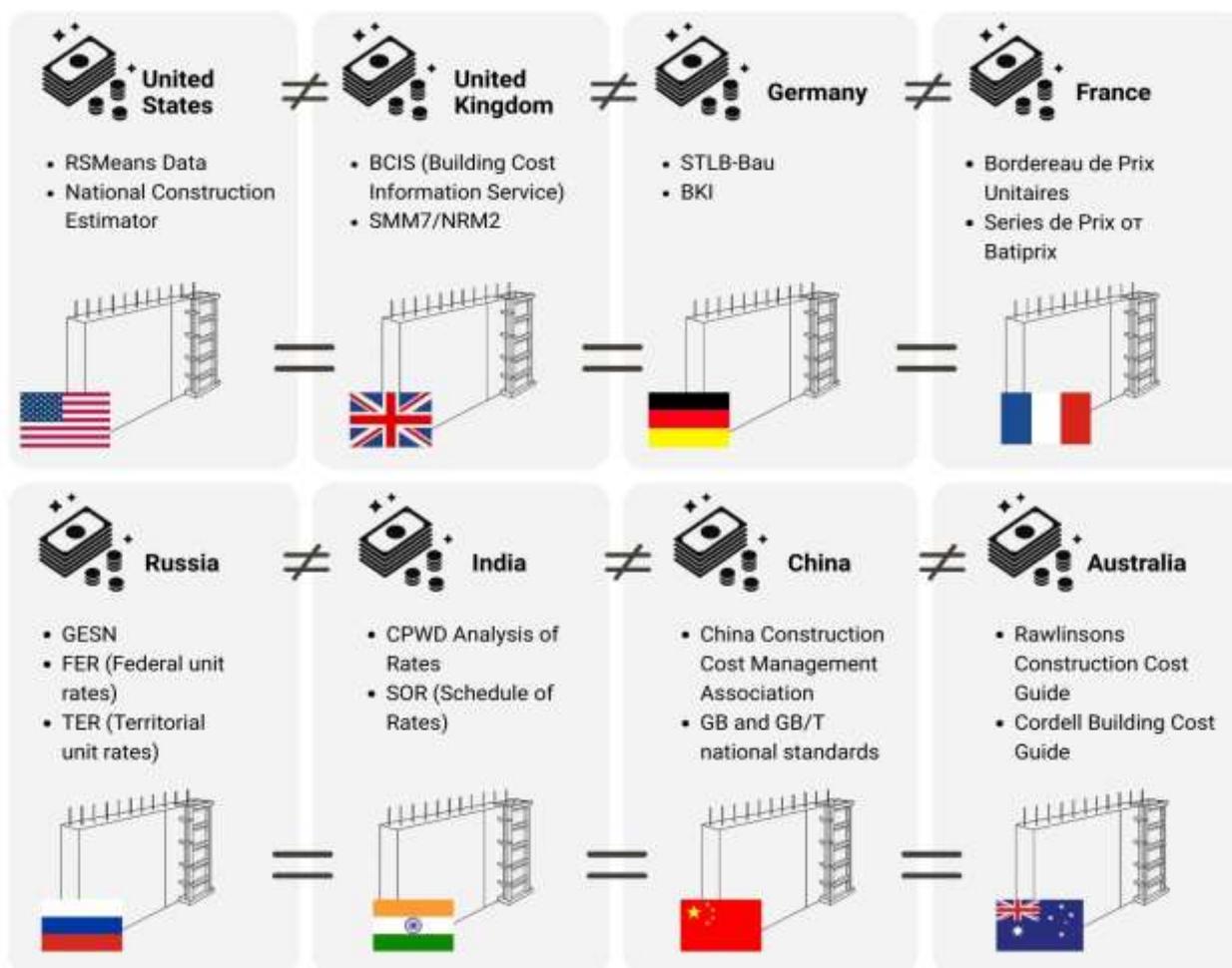
**Фиг. 5.1-6** На етапа на изчисляване на единичната цена на труда се попълват само атриутите на количеството необходими ресурси, всичко останало се зарежда автоматично от базата данни за ресурсите.

Създадените калкулации на работните места се съхраняват под формата на таблици-шаблони на типични проекти, които са пряко свързани с базата данни за строителни ресурси и материали. Тези шаблони представляват стандартизиирани рецепти за повтарящи се видове работа за бъдещи проекти, като осигуряват еднаквост на изчисленията в цялата компания.

Когато разходите за някой ресурс се променят в базата данни (фиг. 5.1-3) - независимо дали ръчно или автоматично чрез изтегляне на текущите пазарни цени (напр. в условия на инфлация) - актуализациите се отразяват незабавно във всички свързани калкулации (фиг. 5.1-6). Това означава, че е необходимо да се променя само ресурсната база, докато шаблоните за калкулиране на разходите и оценките остават непроменени във времето. Този подход осигурява стабилност и възпроизвежданост на изчисленията при всякакви колебания на цените, които се отчитат само в сравнително проста ресурсна таблица (фигура 5.1-3).

За всеки нов проект се създава копие на стандартния шаблон за изчисляване на разходите, което ви позволява да правите промени и да коригирате дейностите, за да отговарят на специфичните изисквания, без да променяте оригиналния шаблон, приет от компанията. Този подход осигурява гъвкавост при адаптирането на изчисленията: можете да вземете предвид спецификата на строителната площадка, желанията на клиента, да въведете коефициенти на рисък или рентабилност (спекулация) - всичко това, без да нарушавате стандартите на компанията. Това помага на компанията да намери баланс между максимизирането на печалбата, удовлетвореността на клиентите и поддържането на своята конкурентоспособност.

В някои страни такива шаблони за остойностяване, трупани в продължение на десетилетия, са стандартизириани на национално ниво и стават част от стандартите на националната система за остойностяване на строителството (фигура 5.1-7).



Фиг. 5.1-7 Различните страни по света имат свои правила за остойностяване със собствени (предписателни) сборници и стандарти за строителни работи за остойностяване на един и същ елемент.

Такива стандартизираны ресурсни бази от оценки (фигура 5.1-7) са задължителни за използване всички участници на пазара, особено за проекти, финансираны с публични средства. Подобна стандартизация осигурява прозрачност, сравнимост и справедливост при формирането на цените и договорните задължения за клиента

## **Окончателно остойностяване на проекта: от оценки до бюджети**

Държавните и специфичните за отрасъла стандарти за оценяване играят различна роля в строителната практика в различните страни. Докато в някои страни се изисква стриктно спазване на един-единствен стандарт, в повечето развити икономики е възприет по-гъвкав подход. В страните с пазарна икономика държавните строителни стандарти обикновено служат само като отправна точка. Строителните компании адаптират тези стандарти към своите модели на работа или изцяло ги преработват, като ги допълват със собствени персонализирани фактори. Тези корекции отразяват корпоративния опит, ефективността на управление на ресурсите и често фактори, в които например могат да бъдат отчетени спекулативните печалби на дружеството.

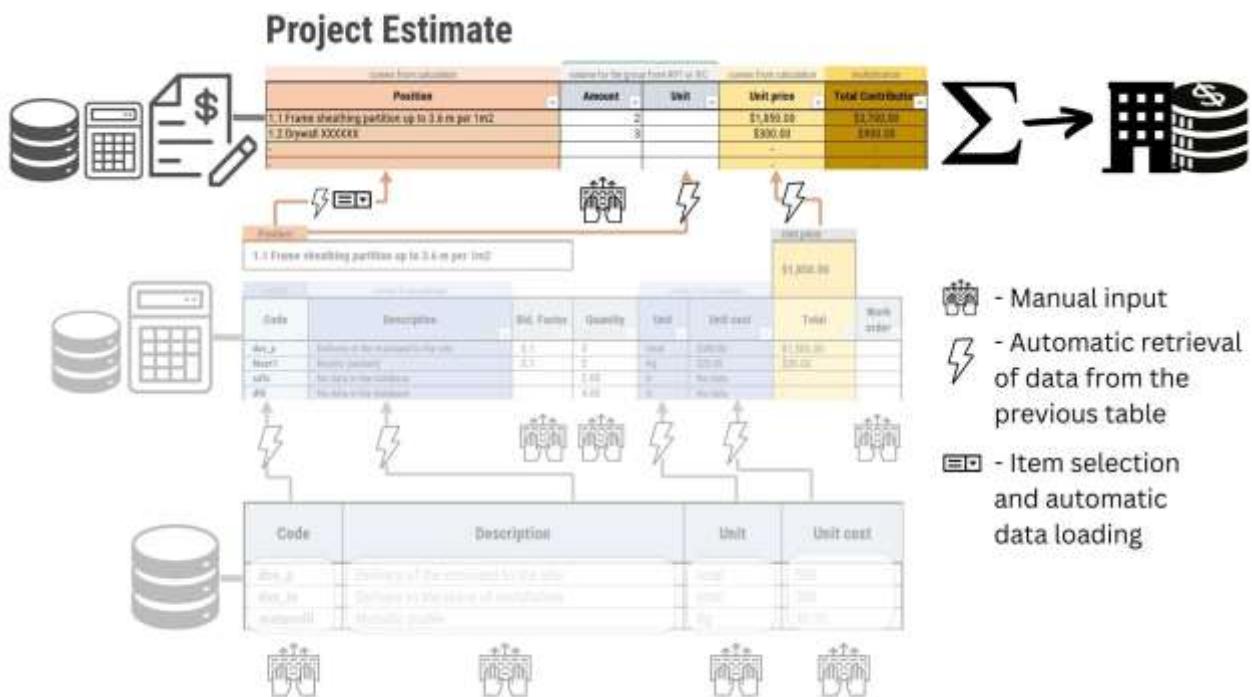
В резултат на това нивото на конкуренция, пазарното търсене, целевите маржове и дори отношенията с конкретни клиенти могат да доведат до значителни отклонения от стандартизираните норми. Тази практика осигурява гъвкавост на пазара, но също така затруднява прозрачното сравняване на оферите на различните изпълнители, като внася елемент на спекулативно ценообразуване в строителния бранш на този етап от изчислителния процес.

След като бъдат подгответи шаблоните за изчисление на отделните дейности и процеси - или, по-често, просто копирани от стандартните правителствени разчети (фиг. 5.1-7) с добавени коефициенти, за да се отразят "особеностите" на конкретната компания - последната стъпка е да се умножат разходите за всяка позиция по съответния атрибут на обхвата на работата или процесите в новия проект.

При изчисляването на общата стойност на проекта за ново строителство ключовата стъпка е да се обобщят разходите за всички елементи на калкулацията, умножени по обема на тези елементи на работата в проекта.

За да създадем общата стойност на проекта, в нашия опростен пример започваме с изчисляване на стойността на изграждането на един квадратен метър стена и умножаваме стойността на нейното изчисление (напр. работата "1m<sup>2</sup> стандартен монтаж на стенни елементи") по общия брой квадратни метри стени в проекта (напр. атрибута "Площ" или "Количество" (фиг. 5.1-8) на единица от тип "Стенни елементи" от CAD на проекта или изчисленията на бригадира).

Изчисляваме разходите за всички елементи на проекта по един и същи начин (фиг. 5.1-8): вземаме разходите за единица работа и ги умножаваме по обема на конкретния елемент или група елементи в проекта. Оценителят трябва само да въведе броя на тези елементи, дейности или процеси в проекта в под формата на обем или количество. Това позволява автоматично да се генерира пълна строителна оценка.



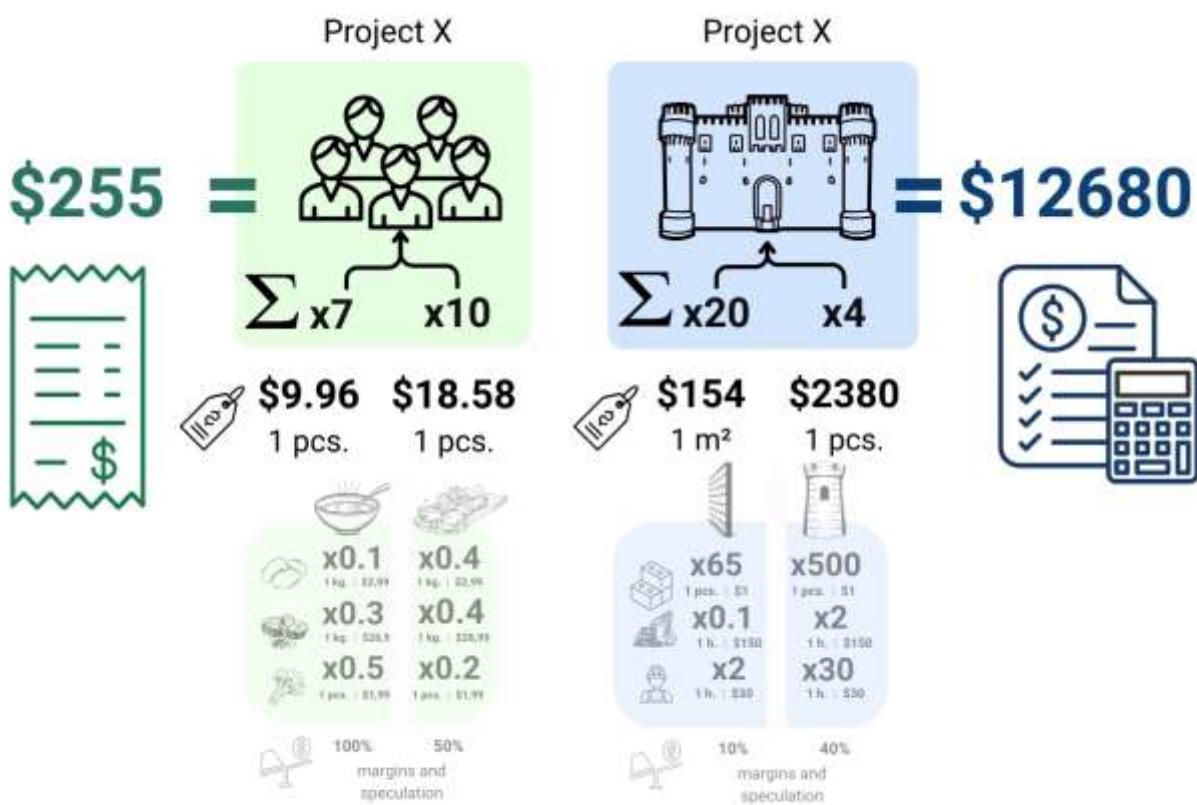
Фиг. 5.1-8 На етапа на създаване на оценка въвеждаме само обхват на работата.

Както и в случая с калкулациите, на това ниво се качват автоматично готови калкулирани елементи (от шаблона на калкулациите или нови, копирани от шаблона и редактирани), които автоматично носят със себе си текущите разходи за единица работа (които се актуализират автоматично от базата данни с ресурси (фиг. 5.1-8 долната таблица)). Съответно, в случай на промяна на данните в ресурсната база данни или таблиците за калкулиране на разходите - данните в калкулацията ще бъдат автоматично актуализирани за текущия ден, без да е необходимо да се променя калкулацията или самата калкулация.

В контекста на ресторант крайната цена на събитието се изчислява по подобен начин и е равна на крайната цена на цялата вечеря, като цената на всяко ястие, умножена по броя на гостите, се прибавя към общата цена на чека (фигура 5.1-9). И точно както в строителството, рецептите за готовене в ресторант може да не се променят в продължение на десетилетия. За разлика от цените, при които цената на съставките може да се променя всеки час.

Точно както собственикът на ресторант умножава цената на всяко ястие по броя на порциите и хората, за да определи общата стойност на събитието, мениджърът по оценка на разходите събира стойността на всички компоненти на проекта, за да получи пълна строителна оценка.

По този начин за всяка дейност в проекта се определя нейната крайна стойност (фиг. 5.1-9), която, умножена по атрибутивния обем на единицата, съответстваща на тази дейност - дава стойността на групите дейности, от които се получава крайната стойност на целия проект.



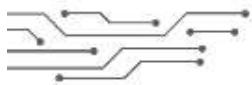
Фигура 5.1-9 Окончателната оценка се изчислява чрез сумиране на атрибута за разходи за работа на всеки елемент по неговия атрибут за обхват.

Общата стойност на проекта (фигура 5.1-8) предоставя финансова картина на проекта, като позволява на клиентите, инвеститорите или финансиращите организации да разберат общия бюджет и финансовите ресурси, необходими за изпълнението на проекта във всеки един ден, като се вземат предвид текущите цени.

И ако процесите на съставяне на ресурсни бази, изчисления и оценки (рецепти на процесите) вече са отработени, полуавтоматизирани и усъвършенствани от десетки хиляди години и записани на държавно ниво, то автоматичното получаване на качествена информация за обема и количеството на елементите за последния етап на окончателната оценка - днес остава тясно място в процесите на всички изчисления на разходните и времевите атрибути на проекта и като цяло на общия бюджет на проекта.

В продължение на хиляди години традиционният метод за изчисляване на обеми е бил ръчният метод за измерване на обеми и количества с помощта на плоски чертежи. С настъпването на цифровата ера компаниите откриха, че информацията за обема и количеството вече може да се извлича автоматично от геометричните данни, съдържащи се в CAD моделите, което революционизира хилядолетните начини за получаване на количествени данни.

Съвременните подходи за оценяване на процесите и прогнозиране включват автоматично извличане на обемни и количествени атрибути от CAD бази данни, които могат да бъдат качени и свързани с процеса на остойностяване, за да се получат актуални обеми на проектните групи на всеки етап от проектирането до експлоатацията.



## ГЛАВА 5.2.

# ИЗЧИСЛЯВАНЕ НА КОЛИЧЕСТВАТА И АВТОМАТИЧНО СЪЗДАВАНЕ НА ОЦЕНКИ И ГРАФИЦИ

## Преминаване от 3D към 4D и 5D: използване на обемни и количествени параметри

С таблиците за изчисляване на разходите с описаните процеси чрез ресурси (фиг. 5.1-8) в ръка, следващата стъпка е автоматично да се получат параметрите за обем или количество за група елементи, които са необходими за изчисленията и за крайната оценка.

Обемните характеристики на елементите на проекта - например стени или площи - могат да се извлечат автоматично от CAD бази данни. Параметричните обекти, създадени в CAD програмите, се преобразуват с помощта на геометрично ядро в числени стойности на дължината, ширината, площта, обема и други параметри. Процесът на получаване на обеми въз основа на 3D геометрия ще бъде разгледан по-подробно в следващата, шеста част (фиг. 6.3-3), посветена на работата с CAD (BIM). В допълнение към обемите, броят на сходните елементи може да се получи и от базата данни на CAD-модела чрез филтриране и групиране на обектите по категории и свойства. Тези параметри, които позволяват групиране, се превръщат в основа за свързване на елементите на проекта чрез изчисленията на ресурсите с калкулациите, окончателните оценки и бюджета на целия проект.

По този начин моделът с данни, извлечен от 3D (CAD) модела, се допълва с нови слоеве с параметри, обозначени като 4D и 5D. В новите слоеве с атрибути на обекти, 4D (време) и 5D (разходи), 3D геометричните данни се използват като източник на стойности на атрибутите на обема на обектите.

- **4D** е слой с информация за параметрите, който добавя информация за продължителността на строителните операции към 3D параметрите на елементите. Тези данни са от съществено значение за планирането на графиците и управлението на сроковете на проекта
- **5D** е следващото ниво на разширяване на модела на данните, при което елементите се допълват с разходни характеристики. По този начин към геометричната информация се добавя финансов аспект: стойността на материалите, работите и оборудването, което дава възможност за изчисляване на бюджета, анализ на рентабилността и управление на разходите по време на строителния процес.

Данните за разходите и атрибутите 3D, 4D и 5D на групите от проектни единици се описват по начин, подобен на изчисленията в модулните ERP, PIMS-системи (или подобни на Excel инструменти), и се използват за автоматично изчисляване на разходите и планиране на бюджета както на отделните групи, така и на целия бюджет на проекта.

## 5D атрибути и получаване на обеми на атрибути от CAD

При изготвянето на окончателната оценка на строителен проект, чието изготвяне разглеждахме в предишните глави (фигура 5.1-8), обемните атрибути за всяка категория елементи на проекта се събират или ръчно, или се извлечат от спецификациите на обемните атрибути, предоставени от

CAD софтуера.

Традиционният ръчен метод за изчисляване на количествата включва анализ на чертежи от бригадира и оценителя, които от хиляди години се представят като линии на хартия, а през последните 30 години - в цифрови формати като PDF (PLT) или DWG. Като се опират на професионалния си опит, те измерват необходимите количества работа и материали, често с линийка и транспортир. Този метод изисква значителни усилия и време, както и специално внимание към детайлите.

Определянето на атрибутите на обхвата по този начин може да отнеме от няколко дни до няколко месеца, в зависимост от обхвата на проекта. Освен това, тъй като всички измервания и изчисления се извършват ръчно, съществува риск от човешка грешка, която може да доведе до неточни данни, което впоследствие се отразява на грешките при оценяването на времето и разходите по проекта, за които ще бъде държана отговорна цялата компания.

Съвременните методи, базирани на използването на CAD бази даннизначително опростяват изчисляването на обемите. В CAD моделите геометрията на елементите вече включва атрибути на обема, които могат да се изчисляват автоматично (чрез геометричното ядро (фиг. 6.3-3)) и да се представлят или експортират в табличен вид.

В такъв случай отделът по оценяване иска от CAD проектанта данни за количествените и обемните характеристики на елементите на проекта. Тези данни се експортират под формата на електронни таблици или се интегрират директно в базите данни за остойностяване - било то Excel, ERP или PMIS - системи. Този процес често не започва с официално искане, а с кратък диалог между клиента (инициатор) и архитекта оценителя от строителната или проектантската компания. По-долу е даден опростен пример, който показва как от ежедневната комуникация се формира структурирана таблица за автоматични изчисления (QTO):

- ❶ Клиент - "Искам да добавя още един етаж към сградата, в същата конфигурация като първия етаж"
- ❷ Архитект (CAD) - "Добавяне на трети етаж, конфигурацията е същата като на втория етаж". И след това съобщение изпраща нова CAD версия на проекта на оценителя.
- ❸ Оценителят автоматично извършва групиране и изчисление (ERP, PMIS, Excel) - "Ще пусна проекта през електронна таблица на Excel с правила за QTO (ERP, PMIS), ще получа обемите по категории за новия етаж и ще генерирам оценка"

В резултат на това текстовият диалог се трансформира в структура на таблица с правила

за групиране:

Елемент	Категория	Под
Припокриване	OST_Floors	3
Колона	OST_StructuralColumns	3
стълбище	OST_Stairs	3

След процеса на автоматично групиране на CAD модела от проектанта в съответствие с правилата на QTO на оценителя и автоматичното умножаване на обемите чрез изчисления на ресурсите (фиг. 5.1-8) се получават следните резултати, които се изпращат на клиента:

Елемент	Обем	Под	Цена за единица.	Общи разходи
Припокриване	420 м <sup>2</sup>	3	150 €/м <sup>2</sup>	63 000 €
Колона	4 бр.	3	2450 €/брой.	9 800 €
стълбище	2 бр.	3	4 300 €/брой.	8 600 €
<b>ОБЩО:</b>	-	-	-	81 400 €

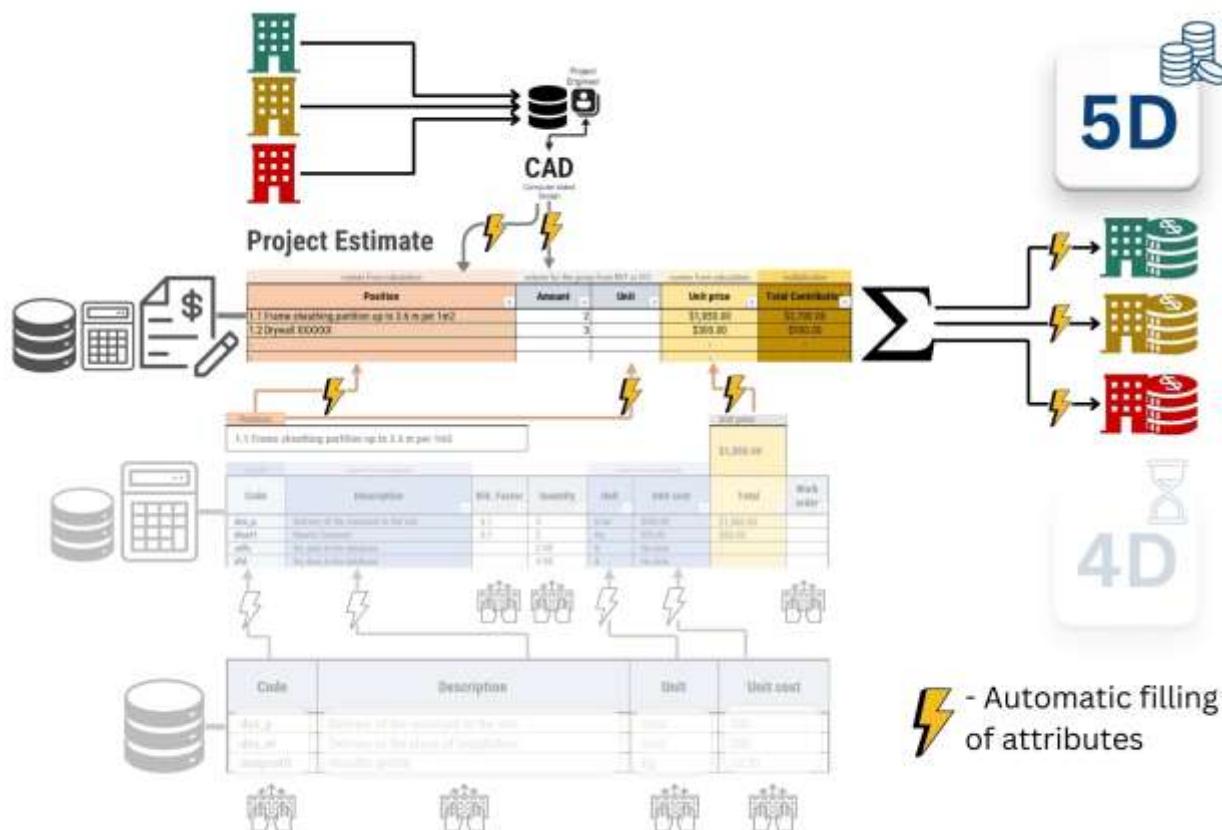
- 💡 Клиентът - "Благодаря ви, това е доста, трябва да съкратим няколко стаи". И цикълът се повтаря многократно.

Този сценарий може да се повтори многократно, особено във фазата на одобрение, когато клиентът очаква незабавна обратна връзка. На практика обаче подобни процеси могат да се проточат с дни или дори седмици. Днес, благодарение на въвеждането на автоматични правила за групиране и изчисление, дейностите, които преди отнемаха значително време, трябва да се извършват за минути. Автоматизираното получаване на количества чрез правила за групиране не само ускорява изчисленията и оценките, но и чрез свеждане до минимум на човешкия фактор намалява вероятността от грешки, осигурявайки прозрачна и точна оценка на разходите по проекта.

Ако при създаването на 3D модела в CAD системата първоначално са били взети предвид изискванията на отдела за оценка (което все още е рядкост в практиката) и имената, идентификаторите на групите елементи и техните класификационни атрибути са зададени под формата на параметри, които съвпадат със структурите на групите и класовете за оценка, тогава обемните атрибути могат да бъдат автоматично прехвърлени към системите за оценка без допълнителни трансформации.

Автоматичното извлечане на обемни атрибути от CAD под формата на таблици със спецификации позволява бързо да се получат актуални данни за разходите за отделните работи и за проекта като

цияло (фиг. 5.2-1). Чрез актуализиране само на CAD файла с обемите на проекта в процеса на изчисление или в системата за изчисление, компанията може бързо да преизчисли оценката, като вземе предвид последните промени, осигурявайки висока точност и последователност на всички последващи изчисления.



Фиг. 5.2-1 Атрибутите на обема от CAD таблици или бази данни се въвеждат автоматично в оценката, което ви позволява незабавно да изчислите общата стойност на проекта.

С нарастващата сложност на капиталовите проекти изчисляването на пълния бюджет и анализирането на общите разходи по проектите при такъв сценарий (фигура 5.2-1) - се превръща в ключов инструмент за вземане на информирани решения.

Според проучването на Accenture "Създаване на по-голяма стойност чрез капиталови проекти (2024 г.)" [20] водещите компании активно интегрират анализа на данни в цифровите инициативи, като използват историческа информация за прогнозиране и оптимизиране на резултатите. Изследването показва, че все повече собственици-оператори прилагат анализ на големи обеми данни, за да прогнозират пазарните тенденции и да оценяват търговската жизнеспособност преди началото на проектирането. Това се постига чрез анализ на складове с данни от съществуващо портфолио от проекти. Освен това 79% от операторите-собственици прилагат "надеждни" прогнозни анализи за оценка на резултатите от проектите и в подкрепа на вземането на оперативни решения в реално време.

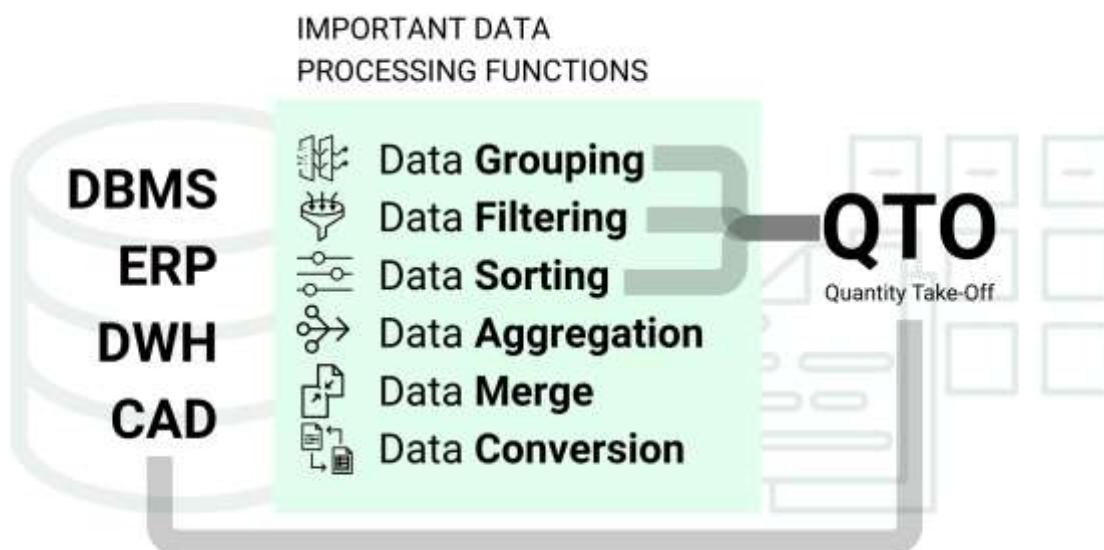
Съвременното ефективно управление на строителни проекти е неразрывно свързано с обработката и анализа на големи количества информация на всички етапи на проектирането и на тези процеси, които го предхождат. Използването на хранилища на данни, ресурсни изчисления, прогнозни модели и машинно обучение позволява не само да се минимизират рисковете при изчисленията, но и да се вземат стратегически решения за финансиране на проекта на ранните етапи на проектирането. Ще говорим повече за складовете за данни и прогнозните модели, които ще допълнят изчисленията, в деветата част на книгата.

Автоматичното получаване на обемните параметри на елементите от CAD проектите, които са необходими за изготвянето на оценките, се извършва с помощта на инструментите за групиране QTO (Quantity Take-Off). Инструментите QTO работят чрез групиране на всички обекти на проекта по специални идентификатори на елементите или параметри на атрибутите на елементите, като се използват спецификации и таблици, създадени в базата данни на CAD.

### **QTO Quantity Take-Off: групиране на данни за проекта по атрибути**

QTO (Quantity Take-Off) в строителството е процесът на извлечение на количествата елементи, необходими за реализацията на даден проект. На практика QTO често е полуручен процес, включващ събиране на данни от различни източници: PDF документи, DWG чертежи и цифрови CAD модели.

Когато се работи с данни, извлечени от бази данни на CAD, процесът QTO се реализира като последователност от операции за филтриране, сортиране, групиране и агрегиране. Елементите на модела се избират според параметрите на класовете, категориите и типовете, след което техните количествени атрибути - като обем, площ, дължина или количество - се обобщават според логиката на изчисленията (фиг. 5.2-2).



**Фиг. 5.2-2 Групирането и филтрирането на данни са най-популярните функции, прилагани в базите данни и складовете за данни.**

Процесът QTO (филтриране и групиране) позволява да се систематизират данните, да се генерират

спецификации и да се подготви входна информация за изчисляване на оценки, покупки и работни графици. Основата на QTO е класификацията на елементите по вида на измерваните атрибути. За всеки елемент или група елементи се избира съответният количествен параметър за измерване. Например:

- **Атрибут за дължина** (бордин камък - в метри)
- **Принадлежност на площта** (работка с гипсокартон - в квадратни метри)
- **Атрибут на обема** (бетонови работи - в кубични метри)
- **Атрибут на количеството** (прозорци - за брой)

В допълнение към обемните характеристики, генериирани математически въз основа на геометрията, при изчисленията след групирането на QTO често се прилагат коефициенти на преразход (фиг. 5.2-12, например 1,1 за отчитане на 10% за логистика и монтаж) - корекционни стойности, които отчитат загубите, монтажа, съхранението или транспорта. Това дава възможност да се предвиди по-точно действителното потребление на материали и да се избегнат както недостигът, така и пре-запасяването на строителната площадка.

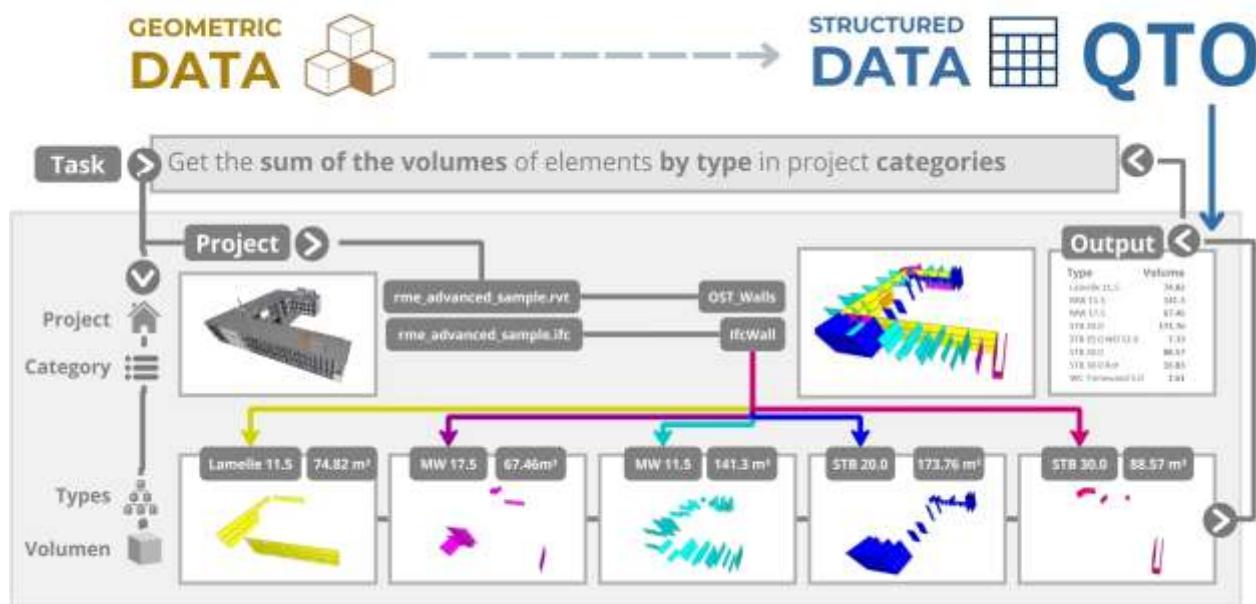
Автоматизираният процес на вземане на количество (QTO) е от съществено значение за изготвянето на точни изчисления и оценки, за намаляване на човешките грешки в процесите на специфициране на обема и за предотвратяване на прекомерното или недостатъчното поръчване на материали.

Като пример за процеса QTO, нека разгледаме често срещан случай, когато е необходимо да се покаже от CAD базата данни таблица-спецификация на обемите по типове елементи за определена категория, класове елементи. Нека да групирате всички елементи на проекта по тип от категорията на стената на CAD проекта и да обобщим атрибутите на обемите за всеки тип, за да представим резултата като QTO таблица на обемите (фиг. 5.2-3).

В примера за типичен CAD проект (фиг. 5.2-3) всички елементи от категорията на стените в CAD базата данни са групирани по тип стена, например "Lamelle 11.5", "MW 11.5" и "STB 20.0", и имат добре дефинирани обемни атрибути, представени в метрични кубове.

Целта на мениджъра, който е на границата между проектантите и специалистите по изчисленията, е да получи автоматизирана таблица с обемите по видове елементи в избраната категория. Не само за конкретен проект, но и в универсална форма, приложима за други проекти с подобна структура на модела. Това позволява подходът да бъде мащабиран и дава възможност за повторно използване на данните без дублиране на усилията.

Отминаха дните, когато опитните проектанти и оценители се въоръжаваха с линийка и внимателно измерваха всяка линия върху хартия или PDF - планове - традиция, която не се е променила през последните хилядолетия. С развитието на 3D - моделирането, при което геометрията на всеки елемент вече е пряко свързана с автоматично изчислени обемни характеристики, процесът на определяне на обемите и количествата QTO стана автоматизиран.



Фигура 5.2-3 Получаването на атрибути на обема и количеството на QTO от даден проект включва групиране и филтриране на елементите на проекта.

В нашия пример задачата е "да се избере категория стени в проекта, да се групират всички елементи по тип и да се представи информация за атрибутите на обема в структуриран, табличен формат", така че тази таблица да може да се използва от десетки други специалисти за изчисляване на разходи, логистика, работни графики и други бизнес казуси (фиг. 6.1-3).

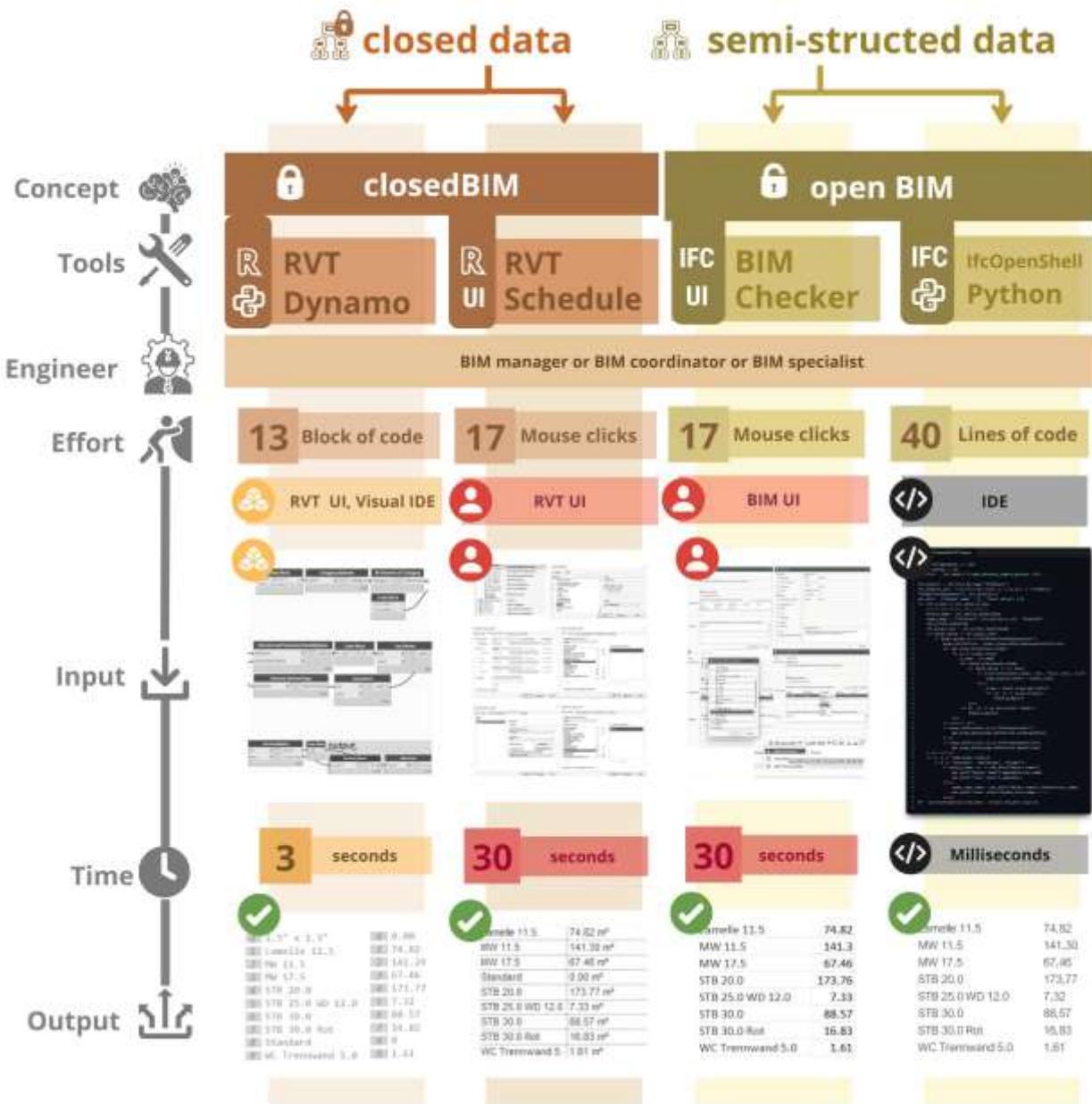
Поради затворения характер на CAD данните днес не всеки специалист може да използва директен достъп до CAD базата данни (прчините и решенията на проблема с достъпа са описани подробно в шестата част на книгата). Поради това много хора трябва да се обърнат към специализирани BIM инструменти, основани на концепциите за отворен BIM и затворен BIM [63]. Когато се работи със специализирани BIM инструменти или директно в средата на CAD софтуера, таблицата с резултатите от QTO (Quantity Take-Off) може да се генерира по различни начини - в зависимост от това дали се използва ръчен интерфейс или софтуерна автоматизация.

Например при използване на потребителския интерфейс на CAD (BIM) софтуер е достатъчно да се извършат около 17 действия (кликования на бутони), за да се получи готова таблица с обеми (фиг. 5.2-4). Въпреки това потребителят трябва добре да разбира структурата на модела и функциите на CAD (BIM) софтуера.

Ако автоматизацията се прилага чрез програмен код или чрез приставки и API инструменти в CAD програми, броят на ръчните стъпки за получаване на обемните таблици се намалява, но е необходимо да се напишат от 40 до 150 реда код, в зависимост от използваната библиотека или инструмент:

- **IfcOpSh (отворен BIM)** или **Dynamo IronPython (затворен BIM)** - позволяват да получите QTO таблица от CAD формат или CAD програма само с ~40 реда код.
- **IFC\_js (отворен BIM)** - изисква около 150 реда код за извлечане на обемни атрибути от модела IFC.

■ **Интерфейс с CAD инструменти (BIM)** - позволява ви да постигнете същия резултат ръчно, със 17 кликвания на мишката.

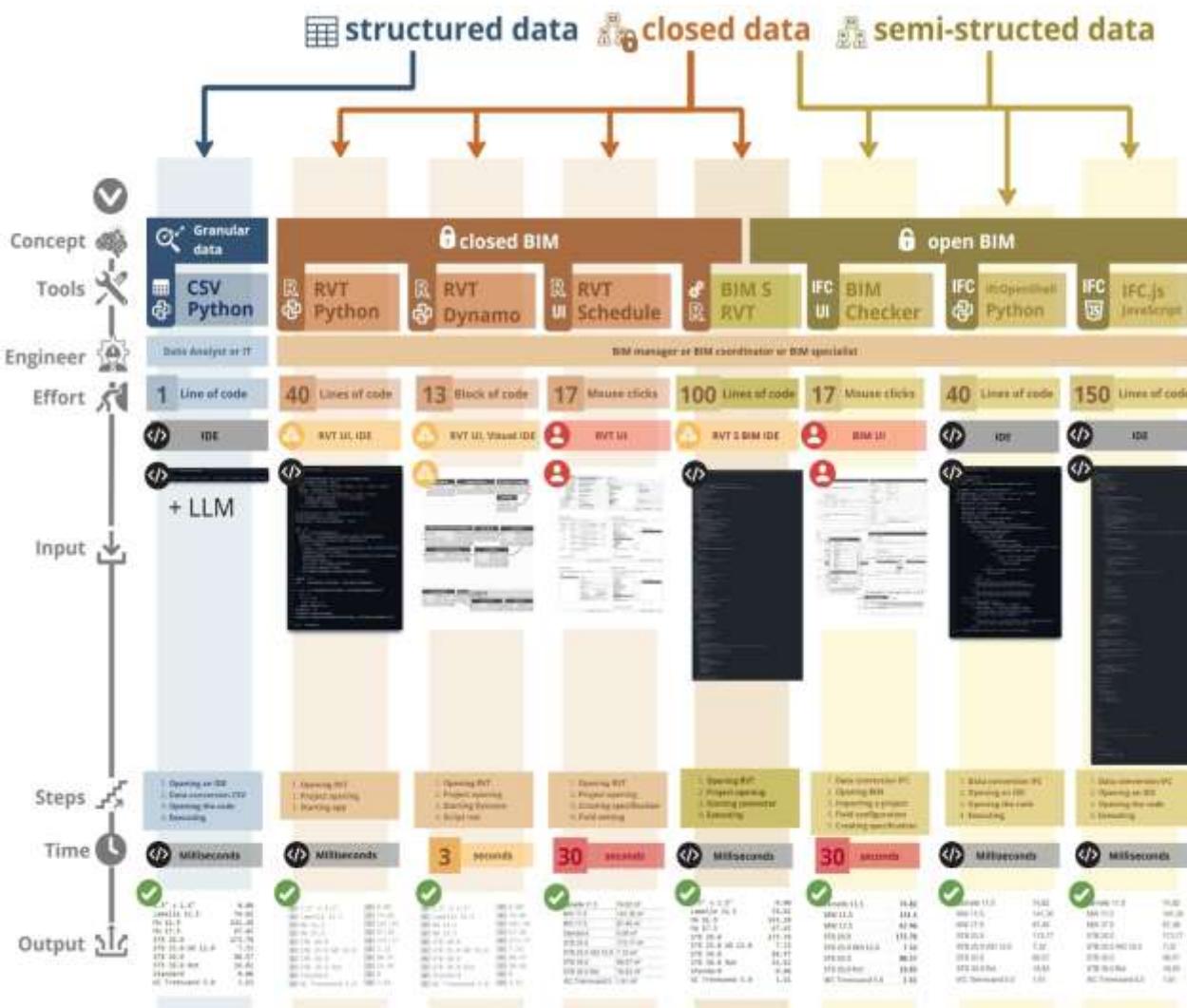


Фигура 5.2-4 CAD (BIM) проектантите и мениджърите, използват от 40 до 150 реда код или десетина натискания на клавиши за създаване на QTO таблици.

Резултатът е същият - структурирана таблица с атрибути за обем за група елементи. Единствената разлика е в цената на труда и нивото на технически познания, необходими на потребителя (фиг. 5.2-4). Съвременните инструменти, по отношение на ръчното събиране на обеми, значително ускоряват процеса на ОТО и намаляват вероятността от грешки. Те позволяват данните да се извлечат директно от модела на проекта, като по този начин се премахва необходимостта от ръчно прерачуващ изчисляване на обемите от чертежите, както се правеше в миналото.

Независимо от използвания метод - дали отворен BIM или затворен BIM - е възможно да се получи идентична QTO - таблица с обемите на елементите на проекта (фиг. 5.2-4). При работа с проектни данни в CAD - (BIM-) концепции обаче потребителите зависят от специализирани инструменти и API, предоставяни от доставчиците (фиг. 3.2-13). Това създава допълнителни нива на зависимост и изисква изучаването на уникални схеми за данни, като същевременно ограничава прекия достъп до данните.

Поради затворения характер на CAD-данныте, получаването на QTO-таблици и други параметри усложнява автоматизацията на изчисленията и интеграцията с външни системи. Чрез използване на инструменти за директен достъп до базите данни и прехвърляне на CAD -данныте на проекта с помощта на инструменти за обратно инженерство в отворен структуриран формат на рамката за данни (фиг. 4.1-13), само с един ред код може да се получи идентична QTO таблица (фиг. 5.2-5 - вариант с гранулирани данни).



Фигура 5.2-5 РАЗЛИЧНИТЕ ИНСТРУМЕНТИ ДАВАТ ЕДНИ И СЪЩИ РЕЗУЛТАТИ ПОД ФОРМАТА НА ТАБЛИЦИ С АТРИБУТИ НА ПРОЕКТНИ ЕДИНИЦИ, Но С РАЗЛИЧНИ РАЗХОДИ ЗА ТРУД.

При използване на отворени структурирани данни от CAD проекти, както е споменато в главата

"Конвертиране на CAD (BIM) данни в структуриран вид", процесът на групиране, QTO, е значително опростен.

Подходите, основани на използването на отворени структурирани данни или на пряк достъп до бази данни с CAD модели, са свободни от маркетинговите ограничения, свързани с акронима BIM. Те разчитат на доказани инструменти, които отдавна се използват в други индустрии (фиг. 7.3-10 ETL процес).

Според проучването на McKinsey "Отворени данни: Разгръщане на иновациите и производителността с помощта на поточна информация" [102], проведено през 2013 г., използването на отворени данни може да създаде възможности за икономии в размер на 30-50 млрд. долара годишно при проектирането, инженеринга, снабдяването и изграждането на съоръжения за производство на електроенергия. Това се изразява в 15% спестяване на капиталови разходи за строителство.

Работата с отворени структурирани (гранулирани) данни опростява извличането и обработката на информация, намалява зависимостта от специализирани BIM платформи и открива пътя към автоматизация, без да е необходимо да се използват патентовани системи или параметрични и сложни модели на данни от CAD формати.

## Автоматизация на QTO с помощта на LLM и структурирани данни

Превеждането на неструктурирани данни в структурирана форма значително подобрява ефективността на различни процеси: то опростява обработката на данните (фиг. 4.1-1, фиг. 4.1-2) и ускорява процеса на валидиране, като прави изискванията на ясни и прозрачни, както вече обсъдихме в предишните глави. По подобен начин преобразуването на CAD данните (BIM) в структурирана отворена форма (фиг. 4.1-12, фиг. 4.1-13) улеснява процеса на групиране на атрибутите и процеса на QTO.

Таблицата с атрибути на QTO има структурирана форма, така че когато използваме структурирани данни от CAD, работим с единен модел на данните (фигура 5.2-5), което премахва необходимостта от преобразуване и превеждане на моделите на данните на проекта и правилата за групиране към общ знаменател. Това ни позволява да групирате данни по един или повече атрибути само с един ред код. За разлика от тях, в отворения BIM и затворения BIM, където данните се съхраняват в полуструктурни, параметрични или затворени формати, обработката изисква десетки или дори стотици редове код и използването на API за взаимодействие с геометрията и атрибутната информация.

- ❷ Пример за групиране на структуриран проект QTO по един атрибут. Текстова заявка в която и да е чат стая на LLM (ChatGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN или друга):

Имам CAD -проект като DataFrame - моля, филтрирайте данните за проекта, за да получите елементи с параметър "Type", съдържащ само стойност "Type 1" ↴

- Много е вероятно отговорът на LLM да бъде под формата на код на Python с помощта на Pandas :



**Фиг. 5.2-6 Един ред код, написан с LLM , ви позволява да групирате цял CAD проект по атрибути "Type" и да получите желаната група елементи.**

Благодарение на простата структура на двуизмерната DataFrame не е необходимо да обясняваме схемата на LLM и модела на данните, което съкращава стъпките на тълкуване и ускорява създаването на крайни решения. По-рано писането дори на прост код изискваше изучаване на езици за програмиране, но сега съвременните езикови модели (LLM) ни позволяват автоматично да превеждаме логиката на процеса в код при работа със структурирани данни с помощта на текстови заявки.

Автоматизацията на LLM и езиковите модели могат напълно да елиминират необходимостта специалистите, работещи с групиране и обработка на данни от CAD (BIM), да изучават езици за програмиране или инструменти за BIM, като предоставят възможност за решаване на проблеми с помощта на текстови заявки.

Същата заявка - групиране на всички елементи на проекта от категорията "стени" и изчисляване на обемите за всеки тип (фиг. 5.2-5) - която в среда на CAD (BIM) изиска 17 кликвания в интерфейса или написване на 40 реда код, в отворени инструменти за обработка на данни (напр. SQL или Pandas) изглежда като пристрастна и интуитивна заявка:

- С един ред в Pandas:

```
df[df['Category'].isin(['OST_Walls'])].groupby('Type')['Volume'].sum()
```

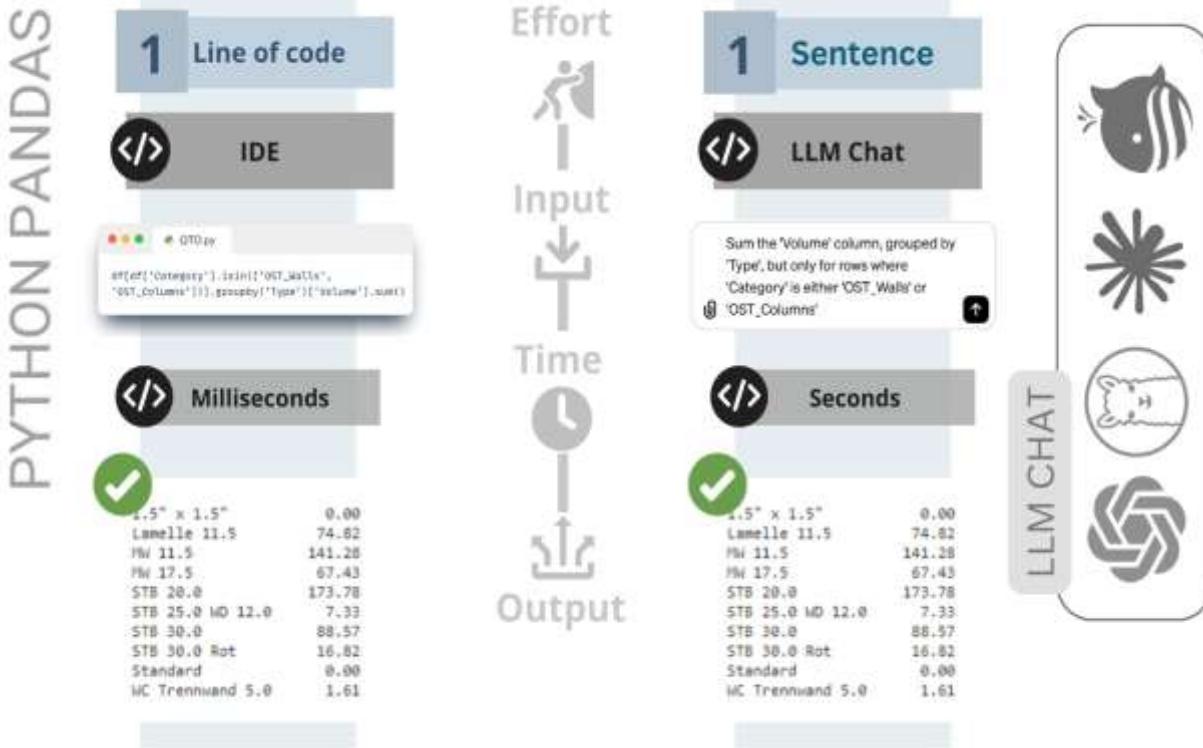
Декодиране на кода: вземете от df (DataFrame) елементите, чиито атрибут-колона "Category", има стойности "OST\_Walls", групирайте всички получени елементи по атрибут-колона "Type" и съберете за получената група елементи атрибути "Volume".

- Групиране на структуриран проект, извлечен от CAD, с помощта на SQL:

```
SELECT Тип, SUM(Обем) AS TotalVolume
FROM элементи
WHERE Категория = 'OST_Walls'
GROUP BY Тип;
```

- С помощта на LLM можем да запишем заявка за групиране в базата данни на проекта като обикновена текстова препратка - подкана (фиг. 5.2-7):

За рамката с данни за проекта групирайте елементите по параметъра "Type", но само за елементите с параметър "Category", равен на "OST\_Walls" или "OST\_Columns", и обобщете параметъра на колоната "Volume" за получената. ↵

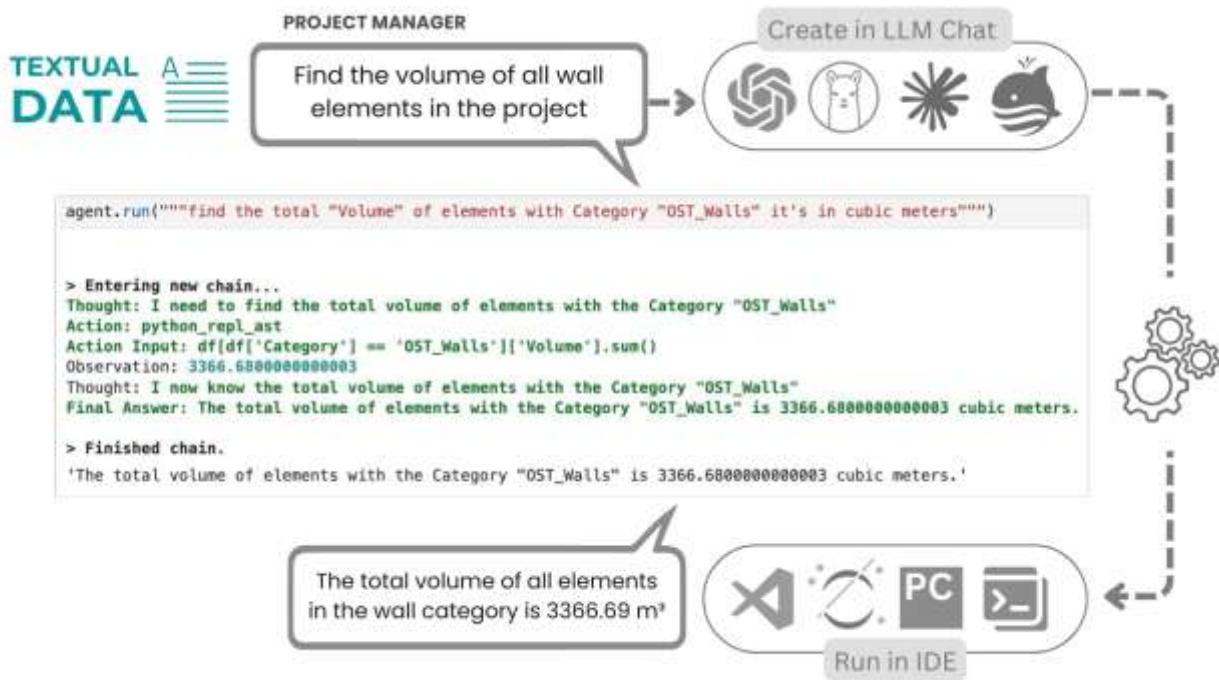


Фигура 5.2-7 С помощта на SQL, Pandas и LLM автоматизацията на обработката на данни вече е възможна с няколко реда код и текстови заявки.

Получаването на QTO от CAD данни с помощта на LLM инструменти (ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok), драстично променя традиционните методи за извличане на атрибутна информация, количествени и обемни данни за отделни обекти и групи обекти.

Сега дори ръководители на проекти, специалисти по остойностяване или логистика, които нямат задълбочени познания в областта на проектирането и не разполагат със специализиран софтуер за CAD - (BIM-) доставчици, имащи достъп до базата данни на CAD, могат да получат общия обем на елементите от категорията на стените или други обекти за няколко секунди, като просто напишат или продиктуват заявка.

При текстовите заявки (фиг. 5.2-8) LLM агентът на модела обработва заявката на потребителя прилагане на определена функция към един или повече параметри - колони на таблицата. В резултат на това потребителят в комуникация с LLM получава или нова колона-параметър с нови стойности, или една конкретна стойност след групиране



**Фиг. 5.2-8** Моделът LLM, работещ със структурирани данни, разбира от контекста на текстовата заявка за какви групи и атрибути пита потребителят.

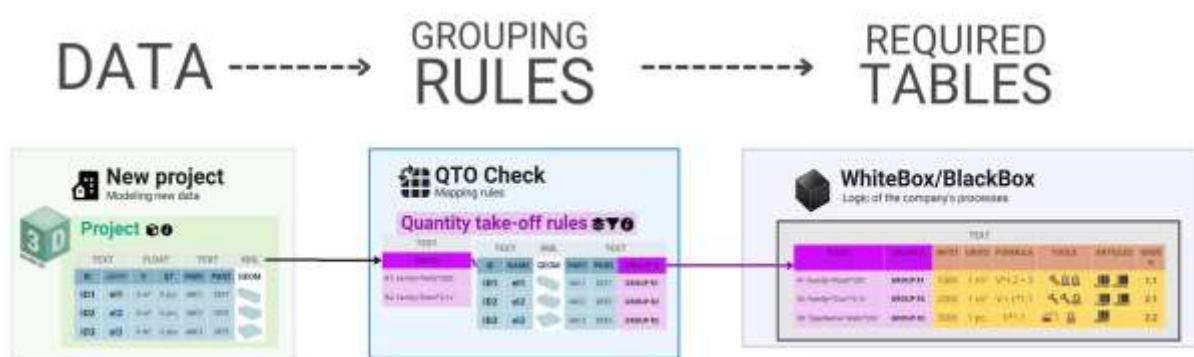
Ако е необходимо да се получат количества само за една група елементи, е достатъчно да се извърши проста QTO заявка (Фигура 5.2-7) върху данните от CAD модела. Когато обаче се изчислява бюджет или оценка за цял проект, състоящ се от много групи елементи, често е необходимо да се извлекат количества за всички видове елементи (класове), като всяка категория елементи се обработва отделно - с групиране по съответните атрибути.

В практиката на оценителите се използват индивидуални правила за групиране и изчисляване на различни видове обекти. Например прозорците обикновено се групират по етажи или зони (параметър за групиране - атрибут Ниво, Стая), а стените - по материал или тип конструкция (параметър Материал, Тип). За да се автоматизира процесът на групиране, такива правила са описани предварително в под формата на таблици с правила за групиране. Тези таблици действат като шаблони

за конфигурация, които определят кои атрибути трябва да се използват при изчисленията за всяка група елементи в проекта.

## QTO изчисление на целия проект с помощта на групови правила от електронна таблица на Excel

В реалните строителни проекти често се налага да се извършва агрегиране по няколко атрибути едновременно в рамките на една група елементи. Например при работа с категорията "Windows" (където атрибутът Category съдържа стойности като OST\_Windows или IfcWindows) елементите могат да бъдат групирани не само по тип - например по стойността в полето Type Name или Type - но и по допълнителни характеристики, като например нивото на топлопроводимост, посочено в съответния атрибут. Това многоизмерно групиране позволява да се получат по-точни резултати за конкретна група. По подобен начин, когато се изчисляват категориите на стените или подовете, като филтри или критерии за групиране могат да се използват произволни комбинации от атрибути - например материал, ниво, под, пожароустойчивост и други параметри (фигура 5.2-9).



Фигура 5.2-9 За всяка група или категория обекти в проекта има различна формула за групиране, състояща се от един или повече критерии.

Процесът на дефиниране на такива правила за групиране е подобен на процеса на създаване на изисквания към данните, описан в главата "Създаване на изисквания и проверка на качеството на данните" (Фигура 4.4-5), където подробно разглеждахме работата с модели на данни. Такива правила за групиране и изчисления гарантират точността и уместността на резултатите за автоматично изчисляване на общите атрибути на количеството или обема на дадена категория обекти, като се вземат предвид всички необходими условия, които трябва да се вземат предвид при изчисленията и изчисленията.

- ❸ Следващата извадка от код филтрира таблицата с проекти, така че полученият набор от данни да съдържа само обекти, в които атрибутната колона "Category" съдържа стойностите "OST\_Windows" или "IfcWindows" и в същото време атрибутната колона "Type" съдържа стойността "Type 1":

Имам проект DataFrame - филтриране на данните, така че в набора от данни да останат само елементите, които имат атрибут "Category", съдържащ стойностите "OST\_Windows" или "IfcWindows", и в същото време атрибутът Type съдържа стойността "Type 1" ↴

💡 Отговор на LLM :

```
df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['Type'].str.contains("Type 1"))]
```

**Фигура 5.2-10 Един ред код, подобен на формулата на Excel , ви позволява да групирате всички обекти на проекта по няколко атрибута.**

Полученият код (фиг. 5.2-10) след превод на CAD данни в структурирани отворени формати (фиг. 4.1-13) може да бъде стартиран в една от популярните IDE (интегрирани среди за разработка), които споменахме по-горе, в онлайн режим: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с PyDev плъгин, Thonny, Wing IDE, IntelliJ IDEA с Python плъгин, JupyterLab или популярни онлайн инструменти: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

💡 За извличане на проектните единици във формуляра QTO DataFrame в категорията "Windows" само с определена стойност на коефициента на топлопроводност можем да използваме следното запитване към LLM:

Имам проект DataFrame - филтриране на данните, така че в набора от данни да останат само записи със стойности "Category", съдържащи "OST\_Windows" или "IfcWindows", и в същото време колоната ThermalConductivity да има стойност 0,. ↴

■ Отговор на LLM :

```
1 df[(df['Category'].isin(['OST_Windows', 'IfcWindows'])) & (df['ThermalConductivity'] == 0.5)]
```

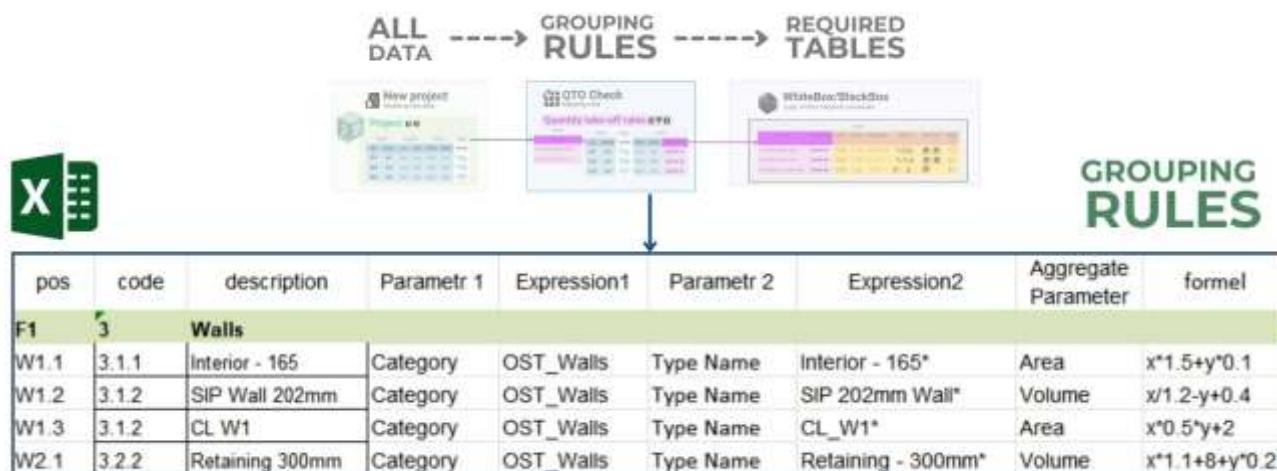
**Фигура 5.2-11** Изключително простият език за заявки Pandas Python ви позволява да изпълнявате КТО за произволен брой проекти едновременно.

В отговора, получен от LLM (фиг. 5.2-11), логическото условие "&" се използва за комбиниране на два критерия: стойност на топлопроводността и принадлежност към една от двете категории. Методът "isin" проверява дали стойността на атрибута-колона "Category" (Категория) се съдържа в предоставения списък.

При проекти с голям брой групи елементи с различна логика на групиране - за всяка категория обекти на проекта (например: прозорци, врати, плочи) трябва да се установят индивидуални правила за групиране, които могат да включват допълнителни коефициенти или формули за изчисляване на общия атрибут. Тези формули (фигура 5.2-12 атрибут "формел", напр. x-стойност на количеството и у-обем на групата) и коефициенти отчитат уникалните характеристики на всяка група, напр:

- % добавки към обема на материала за отчитане на превишенията.
- фиксирано допълнително количество материал
- корекции, свързани с възможни рискове и грешки в изчисленията във формулите

След като правилата за филтриране и групиране са формулирани под формата на формули за параметри за всяка категория елементи, те могат да бъдат съхранени като таблица по редове - например във формат Excel (фигура 5.2-12). Чрез съхраняването на тези правила в структурирана форма процесът на извлечане, филтриране и групиране на данни за проекта може да бъде напълно автоматизиран. Вместо ръчно да се пишат много отделни заявки, системата просто прочита таблицата с параметри и прилага съответните правила към модела (общата рамка с данни за проекта (фигура 4.1-13)), като генерира окончателни таблици за КТО за всяка категория елементи на проекта.



**Фигура 5.2-12 Таблица за групиране на атрибутите на QTO установява правила за групиране на елементите на проекта, като осигурява точен общ брой и обхват за всяка категория.**

Събранныте правила ще ви позволят да групирате целия проект и да извършите всички необходими изчисления, включително корекции на атрибутите на обема. В резултат на това обемите се привеждат в "реален обем", който се използва за изчисления и калкулации, а не тези, които първоначално са били на етап проектиране в CAD модела.

В процеса на автоматично създаване на таблици за обема на QTO за целия проект приложението трябва да премине през всички категории на таблицата с правила за групиране, да вземе атрибутите за групиране, да групира всички елементи на проекта по тях и да обобщи атрибута за обем за тази група, като допълнително го умножи по уточняващ коефициент или фактор.

Нека помолим LLM да ни напише код за такова решение, при което кодът ще трябва да зареди две таблици - таблица с правила за групиране (фигура 5.2-12) и таблица с данни от самия проект (фигура 4.1-13), след което да приложи правилата за групиране, да групира елементите според зададените правила, да изчисли обобщените стойности и да запише резултатите в нов файл на Excel.

💡 Изпратете текстова заявка до чата на LLM:

Нуждае се от код, който да чете данните за проекта от файла 'basic\_sample\_project.xlsx' и след това правилата от 'Grouping\_rules\_QTO.xlsx', чрез които да се групират всички данни по 'Parameter 1' и 'Parameter 2', да се агрегират 'Aggregate Parameter', да се филтрират по 'Expression2', да се извършват изчисления от 'Formel1' и да се записва QTO таблицата в 'QTQ\_table2.xlsx'.

 Отговор на LLM :



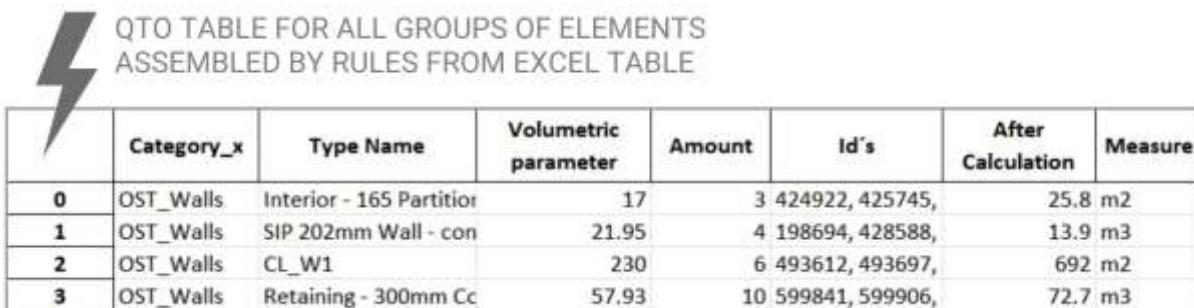
```

1 import pandas as pd
2
3 # Reading CSV and Excel files with project data and grouping rules respectively
4 df = pd.read_csv('../input/rac_basic_sample_project.xlsx', low_memory=False)
5 df_qto = pd.read_excel('../input//Grouping_rules_QTO.xlsx', header=0, skiprows=[1])
6
7 # Cleaning numerical columns in the dataframe by removing non-numeric characters
8 numeric_cols = ['Area', 'Volume', 'Depth', 'Width', 'Length', 'Perimeter']
9 df[numeric_cols] = df[numeric_cols]
10 ].apply(lambda col: pd.to_numeric(col.astype(str).str.extract('(\d+\.\d*)')[0], errors='coerce'))
11
12 # Define a function to process each row in the grouping rules dataframe
13 def process_row(row):
14     # Group by specified parameters and calculate sum and count for the aggregate parameter
15     group = df.groupby([row["Parameter 1"], row["Parameter 2"]])
16     [][row["Aggregate Parameter"]].agg(['sum', 'count'])
17     group = group[group.index.get_level_values(1).str.contains(row["Expression2"], regex=True)]
18
19     # Perform calculation defined in the rules
20     group['After_Calculation'] = eval(row["formel"], {'x': group['sum'], 'y': group['count']})
21
22     return group.reset_index()
23
24 # Apply the function to each row in the grouping rules dataframe and concatenate the results
25 dfgroups = pd.concat([process_row(row) for _, row in df_qto.iterrows()]).reset_index(drop=True)
26
27 # Save the processed data to an Excel file using the OpenPyXL engine
28 dfgroups.to_excel('QTO_table2.xlsx', engine='openpyxl', index=False)

```

Фигура 5.2-13 Код на Python Pandas обработва данни за проектиране от CAD проект, като ги групира в съответствие с правила, определени в Excel .

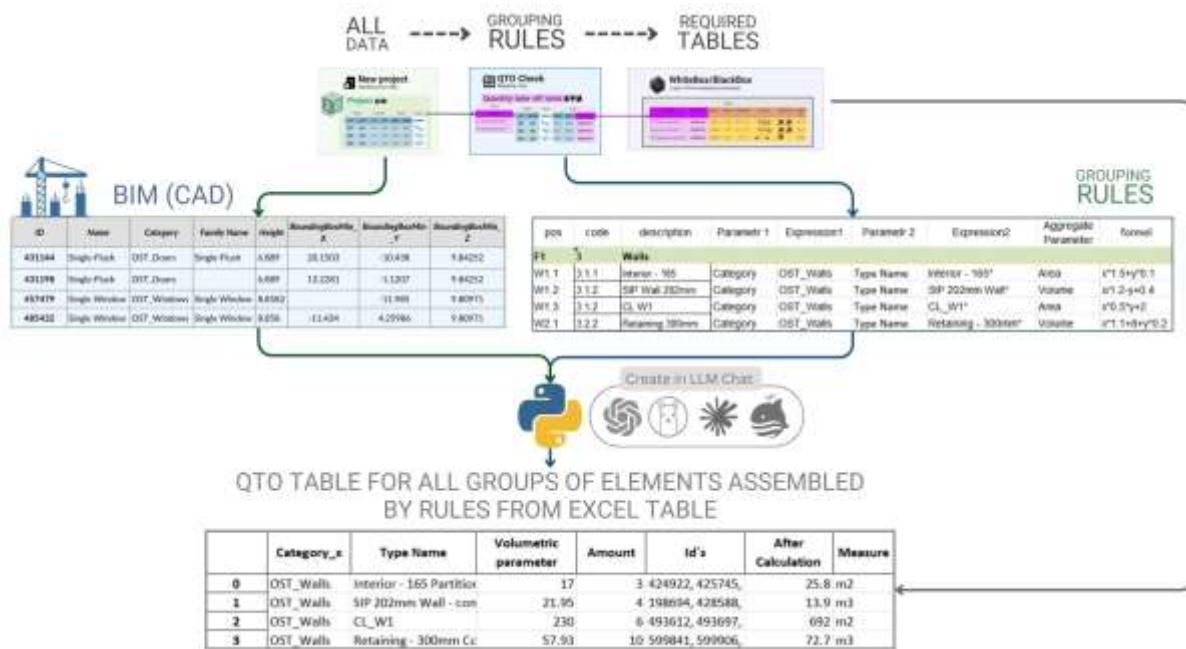
Крайният резултат от изпълнението на кода (фиг. 5.2-13) ще бъде таблица с групи от обекти, която съдържа не само обобщените атрибути на обема от оригиналния CAD- (BIM-) модел, но и нов атрибут на реалния обем, който отчита всички изисквания за правилно създаване на изчисления и оценки (примерна фиг. 5.2-14).



	Category_x	Type Name	Volumetric parameter	Amount	Id's	After Calculation	Measure
0	OST_Walls	Interior - 165 Partition		17	3 424922, 425745,		25.8 m2
1	OST_Walls	SIP 202mm Wall - con		21.95	4 198694, 428588,		13.9 m3
2	OST_Walls	CL_W1		230	6 493612, 493697,		692 m2
3	OST_Walls	Retaining - 300mm Cc		57.93	10 599841, 599906,		72.7 m3

Фигура 5.2-14 Атрибутът "After Calculation" (След изчисление) се добавя към обобщаващата таблица, след като се изпълни код, който автоматично ще изчисли действителния обем.

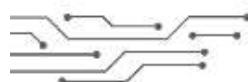
Полученият код (фиг. 5.2-13) може да бъде стартиран в някой от популярните IDE (които споменахме по-горе) и да бъде приложен към произволен брой съществуващи или нови входящи проекти (RVT, IFC, DWG, NWS, DGN и т.н.), било то няколко проекта или може би стотици проекти в различни формати в структуриран вид (фиг. 5.2-15).



**Фигура 5.2-15 Процесът на автоматично групиране на строителни данни свързва BIM данни (CAD) с QTO таблици чрез правила от електронна таблица на Excel.**

Персонализираният и параметризиран процес на събиране на обемни данни (Фигура 5.2-15) позволява напълно автоматизирано събиране на данни за количествените характеристики и обемите на елементите на проекта за по-нататъшна работа с тях, включително оценка на разходите, логистика, работни графици и изчисляване на въглеродния отпечатък и други аналитични задачи.

След като усвоихме инструменти, които ни позволяват лесно да организираме и групирате групи от елементи на проекта по определени признаци, вече сме готови да интегрираме групирани и филтрирани проекти с различни фирмени изчисления и бизнес сценарии.



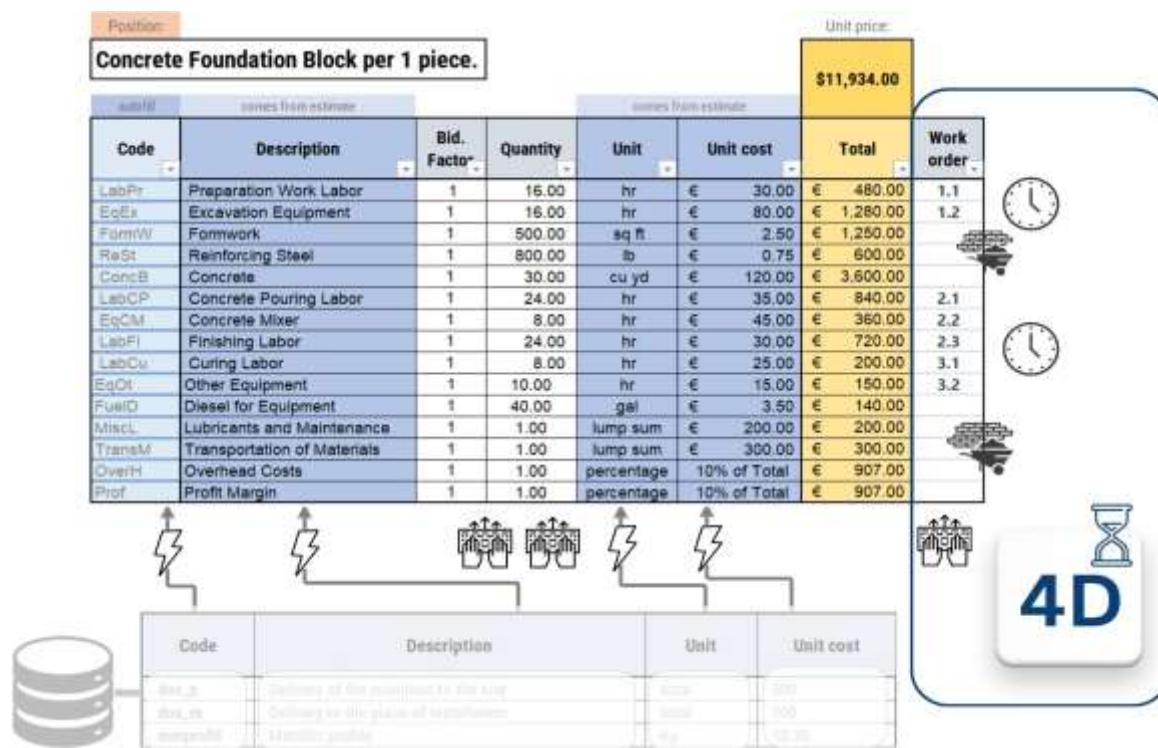
## ГЛАВА 5.3.

### 4Г, 6Г -8Г И ИЗЧИСЛЯВАНЕ НА ЕМИСИИТЕ НА ВЪГЛЕРОДЕН ДИОКСИД CO<sub>2</sub>

#### 4D модел: интегриране на времето в строителните оценки

В допълнение към изчисляването на разходите, едно от ключовите приложения на проектните данни в строителството е определянето на времевите параметри - както за отделните строителни операции, така и за целия проект. Методът за ресурсно базирано оценяване и свързаната с него база данни за изчисления, които бяха разгледани подробно в предишната глава "Изчисления и оценки за строителни проекти", често се използват като основа за автоматизирано изчисляване на времето и създаване на график за изпълнение на работите.

Подходът, основан на ресурсите, отчита не само материалните разходи, но и времевите ресурси. При калкулирането на разходите на всеки процес може да се присвои атрибут на работната поръчка (Фигура 5.3-1 - Параметър на работната поръчка) и количеството време и разходи, свързани с изпълнението на този процес. Тези параметри са особено важни за описание на операции, които нямат фиксирана пазарна цена и не могат да бъдат директно закупени - например използването на строително оборудване, наемането на работници или логистични процеси (които обикновено се изразяват нормално в часове). В такива случаи разходите не се определят от отдела за обществени поръчки, а директно от фирмата-изпълнител въз основа на вътрешни норми или производствени норми (фигура 5.3-1).



**Фиг. 5.3-1 Изчисленията на работата при ресурсно-базирания метод на оценяване включват времето на работните часове.**

По този начин изчисленията на ниво калкулация включват не само разходите за гориво и материали (разходи за закупуване), но и времето, прекарано от шофьорите, техниците и помощните работници на място. В дадения пример (фиг. 5.3-1) таблицата с разходите представлява изчисление на разходите за монтиране на фундаментен блок, включващо съставните етапи на работата, като подготовкa, монтаж на рамката и изливане на бетон, както и необходимите материали и труд. Някои операции обаче, като подготвителните работи, може да нямат разходи за материали, но може да съдържат значителни временни разходи за труд, изразени в човекочасове.

За да се планира последователността на работата (за работния график) на строителната площадка, атрибутът "Работна поръчка" се добавя ръчно към таблицата за изчисления (фиг. 5.3-1). Този атрибут се посочва в допълнителна колона само за елементи, чиято мерна единица е изразена във време (час, ден). Този атрибут е в допълнение към кода на работата, описаните, количеството, мерната единица (параметър "Единица") и разходите. Цифровата последователност (параметър "Work order") на дейностите ви позволява да установите реда, в който се изпълняват задачите на строителната площадка, и да го използвате за планиране.

## График на строителството и неговото автоматизиране въз основа на данни за разходите

Графикът на строителството е визуално представяне на работния план и процесите, които трябва да бъдат извършени в рамките на изпълнението на проекта. Той се създава въз основа на по-

дробни изчисления на ресурсите (фигура 5.3-1), където всяка задача-работка се планира, в допълнение към разходите за ресурси, по време и последователност.

За разлика от усреднените подходи, при които времевите оценки се основават на типичните часове за инсталациране на материали или оборудване, при метода, основан на ресурсите, планирането се основава на действителните данни в калкулацията на разходите. Всяка позиция от оценката, свързана с разходите за труд, се основава на приложения календар, който отчита действителните условия на използване на ресурсите през работния период. Коригирането на производствените часове чрез коефициенти на ниво калкулация (фиг. 5.3-1 параметър "Bid. Factor"), дава възможност да се вземат предвид разликите в производителността и сезонните особености, влияещи върху времето на работа.

За да определим началните и крайните дати на процеса за графика на строителството в диаграмата на Гант, вземаме стойностите на атрибута за количество време за всеки елемент от калкулацията на разходите за фундаментните блокове и ги умножаваме по броя на блоковете (в този случай броя на бетонните фундаментни блокове). Това изчисление дава продължителността на всяка задача. След това нанасяме тези продължителности върху времева линия, започвайки от началната дата на проекта, за да създадем графика и резултатът е визуално представяне, показващо кога трябва да започне и да завърши всяка задача. Параметърът "Работен ред" за процесите допълнително ни позволява да разберем дали работният процес се изпълнява паралелно ("Работен ред", напр. 1.1-1.1) или последователно (1.1-1.2).

**Диаграмата на Гант** е графичен инструмент за планиране и управление на проекти, който представя задачите като хоризонтални ленти върху времева линия. Всяка лента представлява продължителността на задачата, нейното начало и край.

Графикът или диаграмата на Гант помага на ръководителите на проекта и на работниците да разберат ясно кога и в каква последователност трябва да се изпълняват различните етапи на строителството, като се гарантира ефективното използване на ресурсите и спазването на крайните срокове.

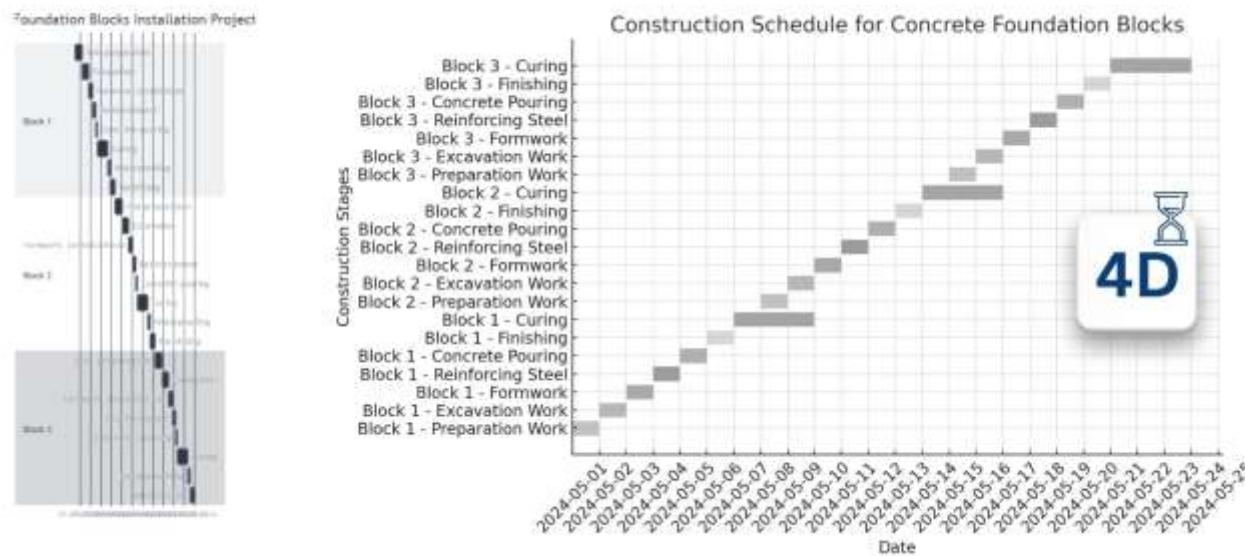
Нека представим график за монтирането на три бетонни блока за основи, като използваме изчисленията от таблицата по-горе. Като използвате таблицата за разходите (фиг. 5.3-1) от примера по-горе, поискайте от LLM да състави график за монтирането на трите елемента на фундаментните блокове, например за първи май 2024 г.

За да изпратим калкулация на разходите в LLM, можем да качим таблицата за калкулация във формат XLSX или просто да вмъкнем еcranна снимка на JPEG изображение на калкулацията директно в чата на LLM (фиг. 5.3-1). LLM самостоятелно ще намери библиотека за визуализиране на изображението на таблицата и като умножи времевите атрибути на дейностите от таблицата по техния обем, ще добави всички данни в графика.

❷ Изпратете текстова заявка до LLM:

Създайте работен график под формата на диаграма на Гант, като използвате стойностите от таблицата, свързана с времето (фиг. 5.3-1 като JPEG в приложението), за монтирането на 3 блока за основи. Работата за всеки блок трябва да се извършва последователно. Началото на работите се определя от 01.05.2024 г. ↴

## 💡 Отговор на LLM

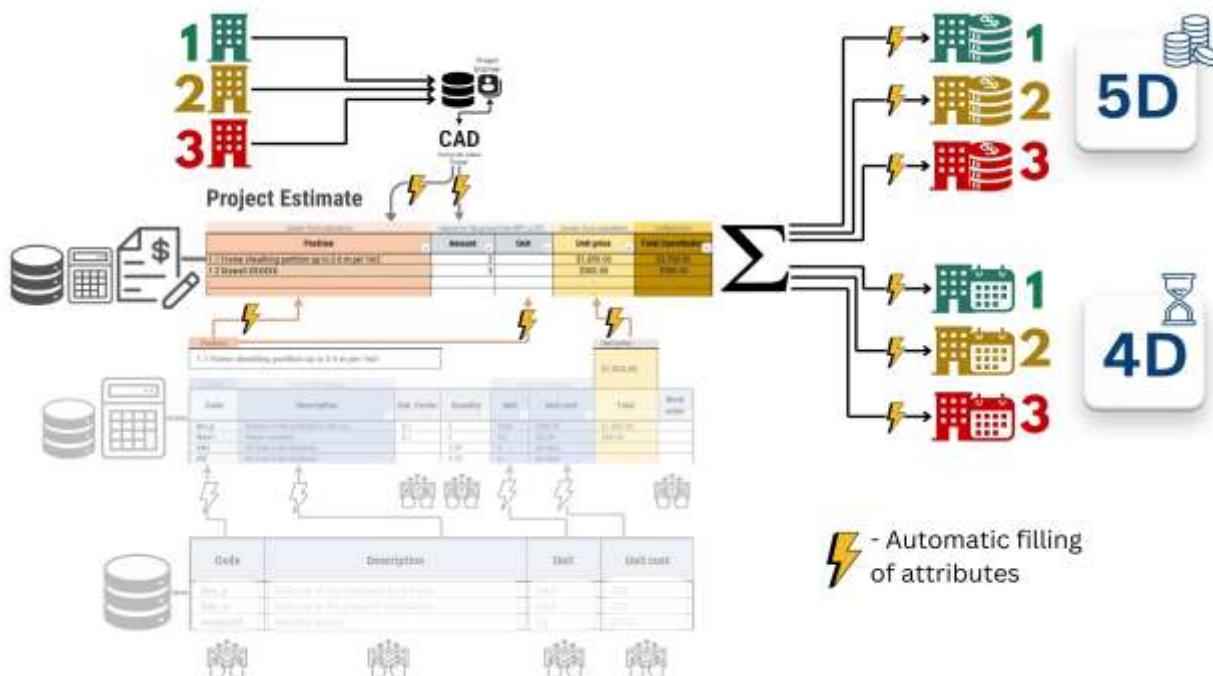


**Фиг. 5.3-2 Автоматично генерирана от няколко LLM Диаграмата на Гант показва етапите на изграждане на три бетонни блока, според условията от подсказката.**

Получената графика (фиг. 5.3-2) представлява времева диаграма, в която всяка хоризонтална лента съответства на определен етап от работата по фундаменталния блок и показва последователността на операциите (параметър "Work order"), като подготовка, изкоп, монтаж на кофраж, армировка, изливане на бетон и довършителни работи, т.е. тези процеси, които са попълнили времевите параметри и последователността в изчисленията.

Такъв график (фиг. 5.3-2) не взема предвид ограниченията, свързани с работните дни, смените или нормите за работно време, а е предназначен единствено за концептуална визуализация на процеса. Точният график, който ще отразява паралелността на работата, може да бъде допълнен с подходящи подсказки или допълнителни инструкции в рамките на чата.

С помощта на единна калкулация на разходите (фиг. 5.3-1), благодарение на обемните атрибути от 3D-геометрията, е възможно автоматично да се оценят както разходите по проекта чрез автоматизирани оценки, така и едновременно с това да се изчислят времевите характеристики на групите под формата на таблици или графики за различните варианти на проекта (фиг. 5.3-3).



Фиг. 5.3-3 Автоматично изчисление, позволява незабавно и автоматично прогнозиране на разходите и времето за различни варианти на проекта.

Съвременните модулни ERP -системи (фиг. 5.4-4), зареждащи данни от CAD модели, използват подобни автоматизирани методи за изчисляване на времето, които значително намаляват процеса на вземане на решения. Това позволява незабавно и точно да се планират работните графици и да се изчислява общото време, необходимо за изпълнение на всички задачи при реализацията на проекта, като се вземат предвид реалните цени.

### Разширени атрибутни слоеве 6D -8D: от енергийна ефективност до осигуряване на безопасност

6D, 7D и 8D са разширени нива на информационно моделиране, всяко от които допринася за допълнителни слоеве от атрибути към цялостния информационен модел на проекта, чиято основа са атрибутите на 3D -модела с техния брой и обхват. Всеки допълнителен слой допринася за специфични параметри, които са необходими за по-нататъшно групиране или по-нататъшна идентификация в други системи, като например системи за управление на собствеността (PMS), компютърно подпомагано управление на съоръженията (CAFM), управление на строителни проекти (CPM) и системи за управление на безопасността (SMS).



**Фигура 5.3-4 Атрибутите 6D, 7D и 8D в информационния модел на данните разширяват разглеждането на различни аспекти на проекта - от енергийната ефективност до безопасността.**

- В **6D**, в допълнение към базата данни на проекта (или рамката с данни (фигура 4.1-13)) с геометрични и обемни атрибути на елементите, се добавя информация (колони с атрибути) за екологичната устойчивост. Това включва информация, свързана с енергийната ефективност, въглеродния отпечатък, възможността за рециклиране на материалите и използването на екологосъобразни технологии. Тези данни позволяват да се оцени въздействието на проекта върху околната среда, да се оптимизират решенията по проекта и да се постигнат целите за устойчиво развитие (ESG).
- **7D** атрибутите допълват атрибутите, необходими за управление на поддръжката на сградите. Това са данни за графиците за поддръжка, жизнените цикли на компонентите, техническата документация и историята на ремонтите. Този набор от информация гарантира, че моделът може да бъде интегриран със системите за поддръжка (CAFM, AMS), позволява ефективно планиране на поддръжката, подмяна на оборудването и осигурява поддръжка през целия жизнен цикъл на обекта.
- **8D** допълнителен атрибутивен слой, - включва информация, свързана с безопасността - както на етапа на строителство, така и по време на последващата експлоатация. Моделът включва мерки за безопасност на персонала, инструкции за действие при аварийни ситуации, системи за евакуация и изисквания за противопожарна защита. Интегрирането на тези данни в цифровия модел помага за предварително отчитане на рисковете и за разработване на архитектурни, инженерни и организационни решения, които отчитат изискванията за здраве и безопасност.

В структурирана таблична форма слоевете от 4D до 8D представляват допълнителни атрибути под формата на колони с попълнени стойности (фигура 5.3-5), добавени към вече попълнените атрибути на 3D - атрибути на модела, като име, категория, тип и обемни характеристики. Стойностите в атрибутните слоеве 6D, 7D и 8D съдържат допълнителни текстови и цифрови данни, като процент на рециклиране, въглероден отпечатък, гаранционен срок, цикъл на подмяна, дата на инсталиране, протоколи за безопасност и др.

The diagram illustrates the evolution of a building information model (BIM) from 3D to 8D. Each stage is represented by an icon and a corresponding table row:

- 3D:** Represented by a construction site icon. The table row shows basic geometric attributes: ID (W-NEW), Type Name (Window), Width (120 cm), Length (-), Recyclability (90%), Carbon Footprint (1622 kgCO<sub>2</sub>e), Warranty Period (8 years), Replacement Cycle (20 years), Maintenance Schedule (Annual), Installation Date (mon), Wellbeing Factors (XYZ Windows), and Safety Protocols (ISO 45001).
- 6D:** Represented by a lightbulb and house icon. The table row adds environmental and energy-related attributes: ID (W-OLD1), Type Name (Window), Width (100 cm), Length (140 cm), Recyclability (90%), Carbon Footprint (1522 kgCO<sub>2</sub>e), Warranty Period (8 years), Replacement Cycle (15 years), Maintenance Schedule (Biannual), Installation Date (08/22/2024), Wellbeing Factors (XYZ Windows), and Safety Protocols (OSHA Standard).
- 7D:** Represented by a building and gear icon. The table row adds operational and maintenance attributes: ID (W-OLD2), Type Name (Window), Width (110 cm), Length (160 cm), Recyclability (90%), Carbon Footprint (1522 kgCO<sub>2</sub>e), Warranty Period (-), Replacement Cycle (15 years), Maintenance Schedule (Biannual), Installation Date (08/24/2024), Wellbeing Factors (????), and Safety Protocols (OSHA Standard).
- 8D:** Represented by a shield with a cross icon. The table row adds safety and installation details: ID (D-122), Type Name (Door), Width (90 cm), Length (210 cm), Recyclability (100%), Carbon Footprint (1322 kgCO<sub>2</sub>e), Warranty Period (15 years), Replacement Cycle (25 years), Maintenance Schedule (Biennial), Installation Date (08/25/2024), Wellbeing Factors (Doors Ltd.), and Safety Protocols (OSHA Standard).

Фигура 5.3-5 6D -8D добавяне на атрибути слоеве към информационния модел на данните, който вече съдържа геометрични и обемни атрибути от 3D -модела.

За нашия нов прозорец (Фигура 4.4-1) елементът с идентификатор W-NEW (Фигура 5.3-5) може да има следните атрибути 3D -8D:

**3D** -атрибути - геометрична информация, получена от CAD системи:

- "Име на типа" - елемент "Window"
- "Ширина" - 120 см
- Освен това можете да добавите точките "Bounding Box" на даден елемент или неговата "геометрия BREP / MESH" като отделен атрибут.

**Атрибути на 6D** - екологична устойчивост:

- Степен на рециклиране от 90 процента
- "Въглероден отпечатък" - 1,622 кг CO<sub>2</sub>

**Атрибути 7D** - данни за управление на обекта:

- "Гаранционен срок" - 8 години
- "Цикълът за подмяна" е на 20 години
- "Поддръжка" - изисква се всяка година

**Атрибути на 8D** - осигуряване на безопасно използване и експлоатация на сградите:

- "Инсталиран" прозорец - от компанията "XYZ Windows"
- "Стандарт за безопасност" - съответства на ISO 45001

Всички параметри, записани в база данни или набор от данни (фиг. 5.3-5), са необходими на специалистите от различни отдели за групиране, търсене или изчисления. Такова многоизмерно описание на обектите на проекта, базирано на атрибути, осигурява пълна картина на техния жизнен цикъл, експлоатационни изисквания и много други аспекти, необходими за проектирането, изграждането и експлоатацията на проекта.

## Оценка на емисиите на CO<sub>2</sub> и изчисляване на емисиите на въглероден диоксид от строителни проекти

Успоредно с темата за устойчивостта на строителните проекти на етап 6D (фигура 5.3-5), съвременното строителство се фокусира върху екологичната устойчивост на проектите, където един от ключовите аспекти става оценката и минимизирането на емисиите на въглероден диоксид CO<sub>2</sub>, които възникват по време на етапите от жизнения цикъл на проекта (например производство и монтаж).

Оценяването и изчисляването на въглеродните емисии на строителните материали е процес, при който общите въглеродни емисии се определят чрез умножаване на обемните характеристики на даден елемент или група елементи, използвани в проекта, по подходящ коефициент на въглеродни емисии за съответната категория.

Разглеждането на въглеродните емисии при оценката на строителни проекти като част от по-широките критерии ESG (екологични, социални и управленски) добавя ново ниво на сложност към анализа. Това е особено важно за клиента-инвеститор при получаването на съответните сертификати, като LEED® (Leadership in Energy and Environmental Design), BREEAM® (Building Research Establishment Environmental Assessment Method) или DGNB® (Deutsche Gesellschaft für Nachhaltiges Bauen). Получаването на един от тези сертификати може значително да повиши пазарната привлекателност на имота, да опости въвеждането в експлоатация и да гарантира спазването на изискванията на наемателите, ориентирани към устойчивото развитие (ESG). В зависимост от изискванията на проекта могат да се използват и сертификатите HQE (Haute Qualité Environnementale, френският стандарт за зелени сгради), WELL (WELL Building Standard, насочен към здравето и комфорта на потребителите) и GRESB (Global Real Estate Sustainability Benchmark).

Околна среда, социална сфера и управление **ESG** (environmental, social and governance) е широк набор от принципи, които могат да се използват за оценка на корпоративното управление, социалното и екологичното въздействие на даден бизнес както във вътрешен, така и във външен план.

ESG, първоначално разработен в началото на 2000 г. от финансови фондове, за да предоставя на инвеститорите информация за широки екологични, социални и управленски критерии, се превърна в ключов показател за оценка както на компании, така и на проекти, включително строителни. Според проучвания на големи консултантски фирми съображенията, свързани с околната среда, социалната сфера и управлението (ESG), се превръщат в неразделна част от строителната индустрия.

Според EY (2023 г.) "Пътят към въглеродна неутралност" компаниите, които активно прилагат ESG -принципите, не само намаляват дългосрочните рискове, но и повишават ефективността на своите бизнес модели, което е особено важно при глобалната трансформация на пазарите [103]. В доклада на PwC "ESG Awareness" се отбелязва, че осведомеността на компаниите за значението на ESG -факторите варира от 67% до 97%, като повечето организации считат тези тенденции за ключови за бъдещата устойчивост [104] и че в по-голямата си част компаниите виждат значителен натиск от страна на заинтересованите страни да интегрират ESG принципите.

По този начин интегрирането на ESG -принципите в строителните проекти не само до-принася за получаването на международни сертификати за устойчивост като LEED, BREEAM, DGNB, но и гарантира дългосрочната устойчивост и конкурентоспособност на компаниите в бранша.

Един от най-значимите фактори, влияещи върху цялостния въглероден отпечатък на даден строителен проект, са етапите на производство и логистика на строителните материали и компоненти. Материалите, използвани на строителната площадка, често оказват решаващо влияние върху общите емисии на CO<sub>2</sub>, особено в ранните етапи от жизнения цикъл на проекта - от добива на сировините до доставката им до строителната площадка.

Изчисляването на емисиите по категории или видове строителни елементи изисква използването на референтни коефициенти за въглеродни емисии, които отразяват количеството CO<sub>2</sub>, генерирано при производството на различни материали. Такива материали са бетон, тухли, рециклирана стомана, алуминий и други. Тези стойности обикновено се извличат отrenomирани източници и международни бази данни, като например UK ICE 2015 (Inventory of Carbon and Energy) и US EPA 2006 (U.S. Environmental Protection Agency) [105]. В следващата таблица (фигура 5.3-6) са обобщени базовите емисионни фактори за редица често срещани строителни материали. За всеки материал са предоставени два ключови параметъра: специфични емисии на CO<sub>2</sub> (в килограми на килограм материал) и коефициенти на преобразуване на обема в теглото (в килограми на кубичен метър), които са необходими за интегриране на изчисленията в модела за проектиране и свързване с групиранието на данните в QTO.



**CO<sub>2</sub>**

Carbon Emitted in Production		UK ICE Database (2015) USEPA (2006)	UK ICE Database (2015) USEPA (2006)	Coefficient m <sup>3</sup> to kg
Material	Abbreviated	Process Emissions (kg CO <sub>2</sub> / kg of product) (X1)	Process Emissions (kg CO <sub>2</sub> / kg of product) (X2)	Kg / m <sup>3</sup> (X3)
Concrete	Concrete	0.12	0.12	2400
Concrete block	Concrete_block	0.13**	0.14	2000
Brick	Brick	0.24	0.32	2000
Medium density fiberboard (MDF)	MDF	0.39*	0.32	700
Recycled steel (avg recy content)	Recycled_steel	0.47	0.81	7850
Glass (not including primary mfg.)	Glass	0.59	0.6	2500
Cement (Portland, masonry)	Cement	0.95	0.97	1440
Aluminum (virgin)	Aluminum	12.79	16.6	2700

**Фигура 5.3-6 Количество въглеродни емисии, отделени при производството на различни строителни материали, според базата данни на ICE на Обединеното кралство и EPA на САЩ.**

За да се изчислят общите емисии на CO<sub>2</sub> за даден проект, както при изчисленията на 4D и 5D, е необходимо да се определят атрибутните обеми на всяка група съоръжения. Това може да се направи с помощта на инструменти за количествен анализ (QTO), като се получат атрибутните обеми в кубични метри, както е разгледано подробно в раздела за извличане на количества. След това тези обеми се умножават по съответните коефициенти за атрибута "Процесни емисии на CO<sub>2</sub>" на всяка група материали.

- Нека автоматично извлечем таблицата с обемите по тип елемент от проекта CAD (BIM), като групирате всички данни за проекта, както вече беше направено в предишните глави. За да изпълним тази задача, нека направим справка с LLM.

Моля, групирайте таблицата DataFrame от проекта CAD (BIM) по параметъра на колоната "Име на обекта" (или "Тип") и покажете броя на елементите във всяка група, както и обобщете параметъра "Обем" за всички елементи от типа. ↴

- Отговор на LLM :

```

Create in LLM Chat
Data
f(x)
Output
Run in IDE
  
```

```

1 import pandas as pd
2
3 # Import Projekt as CSV tain DataFrame
4 df = pd.read_csv('your_dataframe.csv')
5
6 # Group, count, and sum 'Volume' in one line
7 result = df.groupby('ObjectType')['Volume'].agg(Amount='size', Sum of Volumes, m³='sum')
  
```

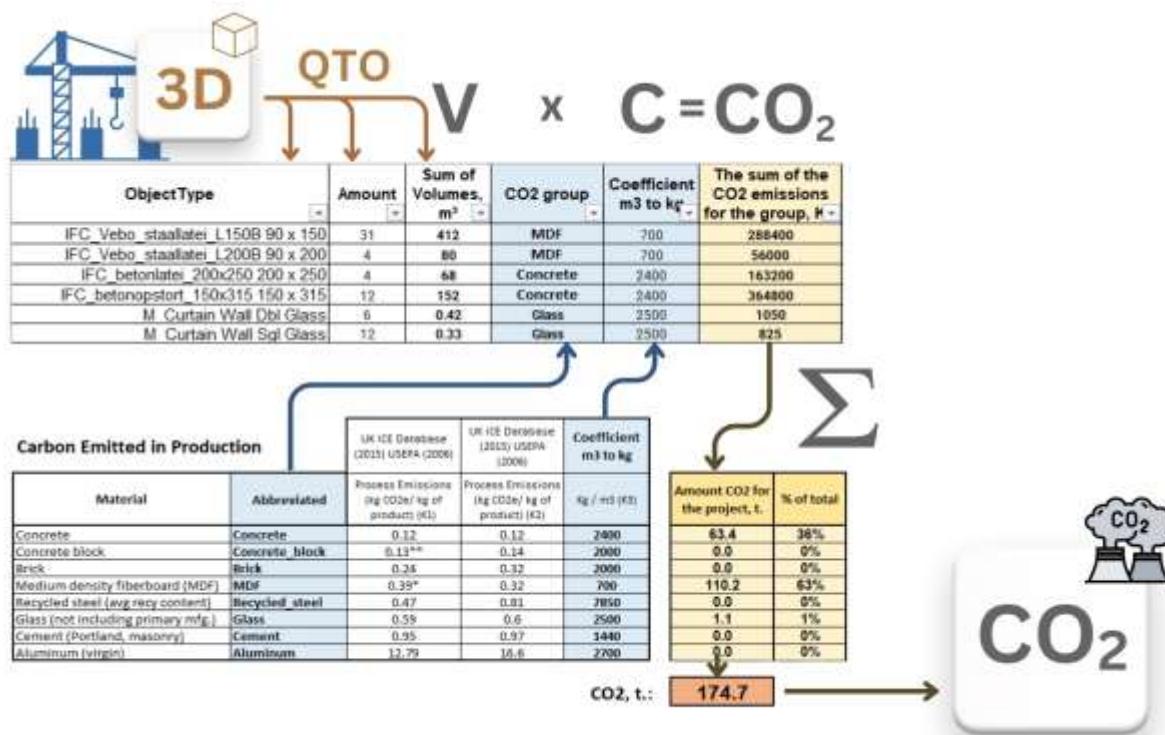
ObjectType	Amount	Sum of Volumes, m <sup>3</sup>
IFC_Vebo_staillei_L150B_90_x_150	31	412
IFC_Vebo_staillei_L200B_90_x_200	4	80
IFC_betonlatei_200x250_200_x_250	4	68
IFC_betonopstort_150x315_150_x_315	12	152
M_Curtain_Wall_Dbl_Glass	6	0.42
M_Curtain_Wall_Sgl_Glass	12	0.33

Фигура 5.3-7 Генерираният код в LLM е групирал за нас обекти на проекта по тип (ObjectType) с обобщен атрибут "Volume".

За да автоматизирате изчисляването на общите емисии на CO<sub>2</sub> за целия проект, просто настройте автоматично съпоставяне на данните в таблицата или ръчно свържете типовете елементи (Фигура 5.3-7) със съответните типове материали (Фигура 5.3-6) от таблицата с емисионните фактори. Готовата таблица с емисионни фактори и формули, както и код за извлечане на обеми от CAD формати (BIM) и автоматизиране на определянето на CO<sub>2</sub> можете да намерите в GitHub, като потърсите "CO<sub>2</sub>\_calculating-the-embodyied-carbon. DataDrivenConstruction." [106].

По този начин интегрирането на данни след групиране на елементите на QTO от базата данни CAD позволява автоматично изчисляване на емисиите на въглероден диоксид (фиг. 5.3-8) за различни варианти на проектиране. Това дава възможност да се анализира въздействието на различните материали в различни варианти и да се изберат само тези решения, които отговарят на изискванията на клиента за емисиите на CO<sub>2</sub>, за да се получи конкретен сертификат при въвеждането на сградата в експлоатация

Оценяването на емисиите на CO<sub>2</sub> чрез умножаване на коефициентите по обемите на групирани елементи на проекта е типичен пример за задача в процеса на получаване от строителна компания на ESG рейтинг (напр. сертификат LEED) за даден обект.

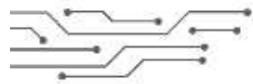


Фигура 5.3-8 Интегрирането на QTO групи от CAD бази данни осигурява точност и автоматизация при извеждането на оценки на крайните емисии на CO<sub>2</sub>.

По подобен начин, чрез дефиниране на обемите на групите елементи, можем да извършваме изчисления за контрол и логистика на материалите, мониторинг и управление на качеството, енергийно моделиране и анализ, както и множество други задачи, за да получим нов статус на атрибута (параметър в таблицата) както за отделните групи елементи, така и за целия проект.

Ако броят на тези изчислителни процеси в компанията започне да нараства, възниква въпросът за необходимостта от автоматизиране на тези изчисления и внедряване на резултатите от тях в процесите и системите за управление на данни на компанията.

Поради сложността на едно цялостно решение средните и големите компании в строителния бранш възлагат такава автоматизация на фирми за разработване на ERP (или PMIS) системи. Развойните компании създават за големите клиенти една цялостна модулна система за управление на много различни информационни слоеве, включително изчисления на материали и ресурси.



## ГЛАВА 5.4.

# СИСТЕМИ ЗА СТРОИТЕЛСТВО ERP И PMIS

### Строителни ERP -системи на примера на изчисления и оценки

Модулните ERP системи интегрират различни атрибутни (информационни) слоеве и потоци от данни в една цялостна система, като позволяват на ръководителите на проекти да управляват ресурсите, финансите, логистиката и други аспекти на проекта по синхронизиран начин в рамките на една платформа. Една строителна ERP система действа като "мозък" на строителните проекти, опростявайки повтарящи се процеси чрез автоматизация, осигурявайки прозрачност и контрол в целия строителен процес.

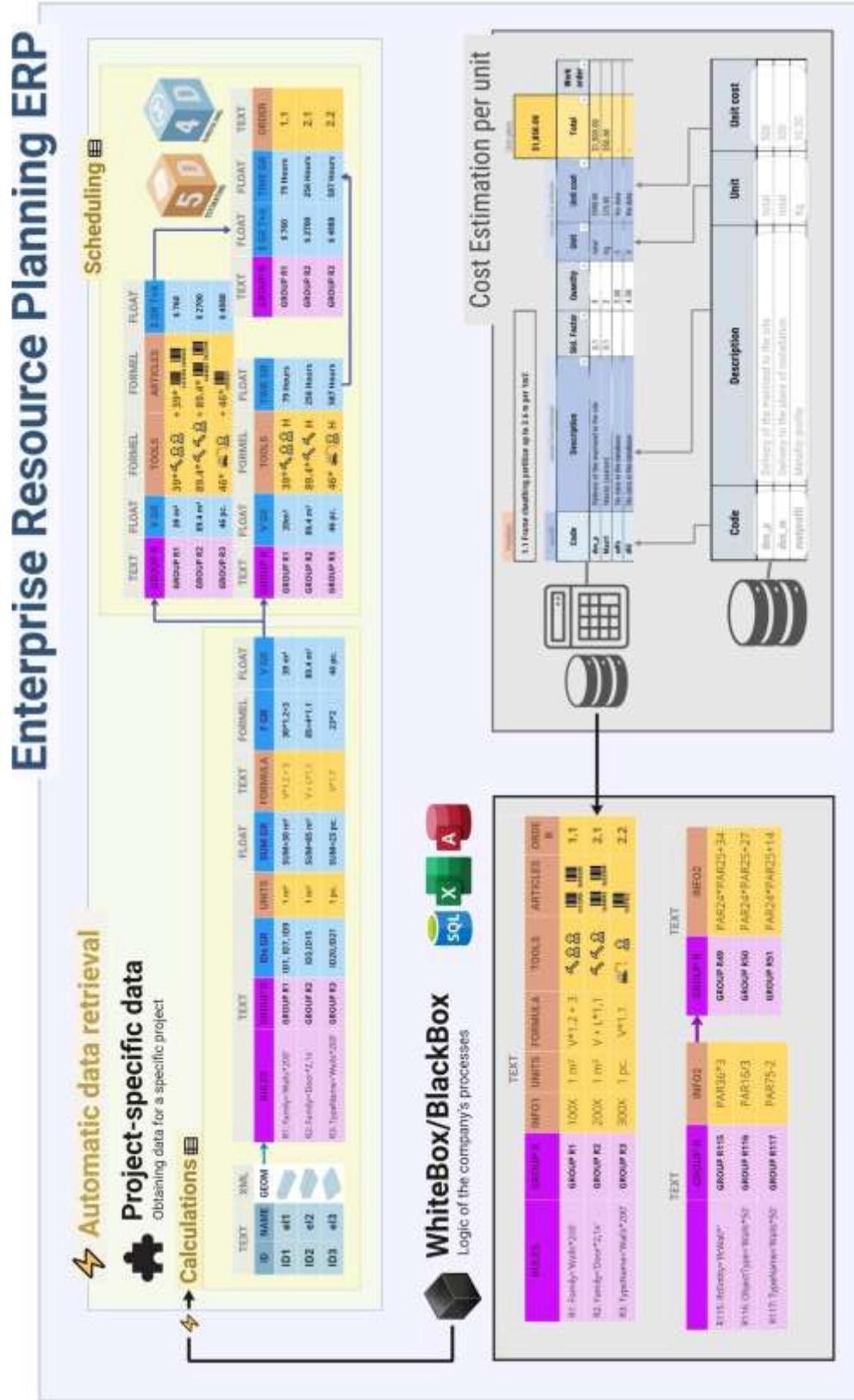
**Строителни ERP -системи** (Enterprise Resource Planning) са цялостни софтуерни решения, предназначени за управление и оптимизиране на различни аспекти на строителния процес. В основата на строителните ERP системи са модулите за управление на калкулирането на разходите и планирането, което ги прави важен инструмент за ефективно планиране на ресурсите.

ERP модули -системите позволяват на потребителите да въвеждат, обработват и анализират данни по структуриран начин, обхващащи различни аспекти на проекта, които могат да включват отчитане на разходите за материали и труд, използване на оборудването, управление на логистиката, човешки ресурси, контакти и други строителни дейности.

Един от функционалните блокове на системата е модулът за автоматизация на бизнес логиката - *BlackBox/WhiteBox*, който играе ролята на център за управление на процеса.

*BlackBox /WhiteBox* позволява на специалистите, използващи ERP -система, да управляват гъвкаво, чрез права за достъп, различни аспекти на бизнеса, които вече са били предварително конфигурирани от други потребители или администратори. В контекста на ERP системите термините *BlackBox* и *WhiteBox* се отнасят до нивата на прозрачност и възможност за контрол на вътрешната логика на системата:

- **BlackBox** ("черна кутия") - потребителят взаимодейства със системата чрез интерфейса, без да има достъп до вътрешната логика на изпълнение на процеса. Системата извършва изчисления самостоятелно, въз основа на предварително определени правила, скрити от крайния потребител. Той въвежда данните и получава резултата, без да знае какви атрибути или кофициенти са използвани вътре.
- **WhiteBox** ("бяла кутия") - логиката на процеса е достъпна за преглед, персонализиране и модифициране. Напредналите потребители, администратори или интегратори могат ръчно да дефинират алгоритми за обработка на данни, правила за изчисления и сценарии за взаимодействие между субектите на проекта.



Фиг. 5.4-1 Архитектура на строителна ERP -система, за получаване на оценки и графики за работа чрез ръчно попълване на обемните атрибути.

Пример за това е, когато опитен потребител или администратор зададе правило: кои атрибути в оценката трябва да се умножат помежду си или да се групират по определен признак и къде да се запише крайният резултат. По-късно по-малко обучените специалисти, като например инженерите по изготвяне на оценки, просто зареждат нови данни в ERP системата чрез потребителския интерфейс - и получават готови оценки, графици или спецификации, без да се налага да пишат код или да разбират техническите подробности на логиката.

В предишните глави изчислителните и логическите модули бяха разгледани в контекста на взаимодействието с LLM. В среда на ERP такива изчисления и трансформации се извършват в модули, скрити зад интерфейс от бутони и форми.

В следващия пример (фиг. 5.4-1) администраторът на ERP -системата в модула BlackBox /WhiteBox е определил правила за съпоставяне на атрибути на обекти от оценки с атрибути за групиране на QTO. Благодарение на това конфигуриране в модула BlackBox/WhiteBox (от мениджъра или администратора) потребителят (оценител или инженер), добавяйки ръчно атрибут за количество или обем чрез потребителския интерфейс на ERP, автоматично получава готовите оценки и работни графици. По този начин процесите на изчисление и генериране на оценки, разгледани в предходните глави, с помощта на код, вътре в ERP, се превръщат в полуавтоматичен конвейер.

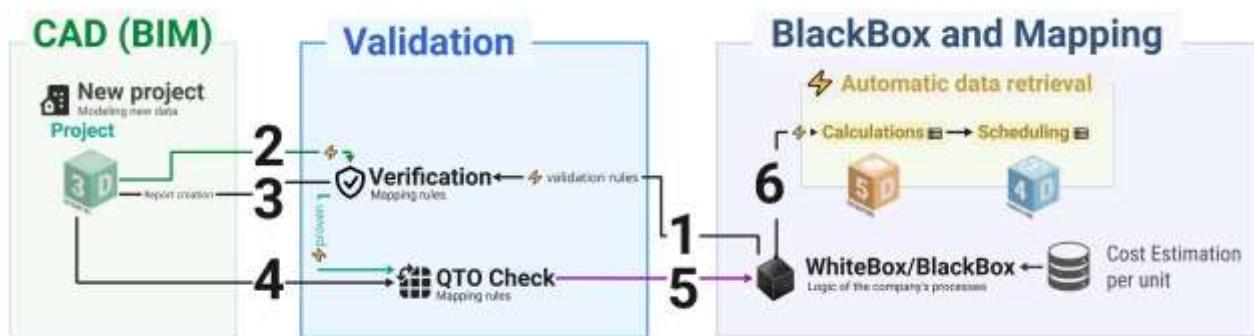
Свързването на този полуавтоматичен процес с обемните атрибути от CAD (BIM) моделите (фиг. 4.1-13), например чрез зареждане на CAD проекта в ERP модула, който е предварително конфигуриран за тази цел, превръща потока от данни в синхронизиран механизъм, способен самостоятелно и незабавно да актуализира стойността на отделните групи елементи или на целия проект в отговор на всички промени в него по време на фазата на проектиране, при зареждане на CAD модела в ERP.

За да се създаде автоматизиран поток от данни (фиг. 5.4-2) между CAD (BIM) и ERP системите, основните процеси и изисквания към данните от базите данни с CAD (BIM) модели трябва да бъдат определени по структуриран начин, както вече обсъдихме в горната глава "Изисквания и осигуряване на качеството на данните". Този процес в ERP системите е разделен на подобни стъпки:

- **Създаване на правила за валидиране (1)**, които играят важна роля за осигуряване на точността на данните, постъпващи в ERP -системата. Правилата за валидиране служат като филтри, които валидират субектите и техните атрибути, като позволяват само на тези елементи, които отговарят на изискванията, да влязат в системата. Можете да научите повече за проверката и валидирането в главата "Създаване на изисквания и валидиране на качеството на данните".
- След това в ERP **се извършва процес на проверка (2)**, който потвърждава, че всички елементи на проекта с техните атрибути и стойности са създадени правилно и са готови за следващите стъпки на обработка.
- Ако има проблеми с непълни данни за атрибутите, **се генерира доклад (3)** и проектът, заедно с инструкциите за корекция, се изпраща за преразглеждане, докато стане готов за следващата итерация.
- След като данните за проекта бъдат валидирани и проверени, те се използват в друг модул на ERP **(4) за създаване на таблици за количеството (QTO)**, които създават атрибути

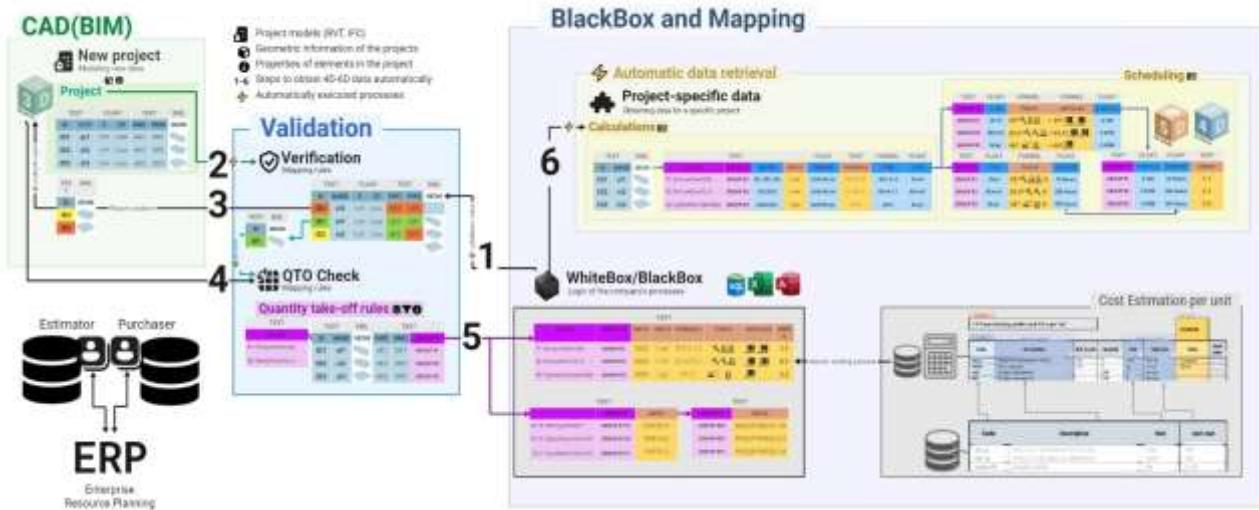
за количеството за групите субекти, материалите и ресурсите в съответствие с предварително създадени правила (WhiteBox/BlackBox).

- Групираните данни по правила за съвпадение или QTO **се интегрират** автоматично **с изчисления** (напр. разходи и време) (5).
- В последната стъпка на ERP -системата потребителят, като умножава атриутите на обхвата от таблицата QTO с атриутите на таблиците на процесите (напр. оценявани елементи), **автоматично генерира резултати от изчисленията** (6) (напр. оценки на разходите, работни графици или емисии на CO<sub>2</sub>) за всяка група субекти и за проекта като цяло.

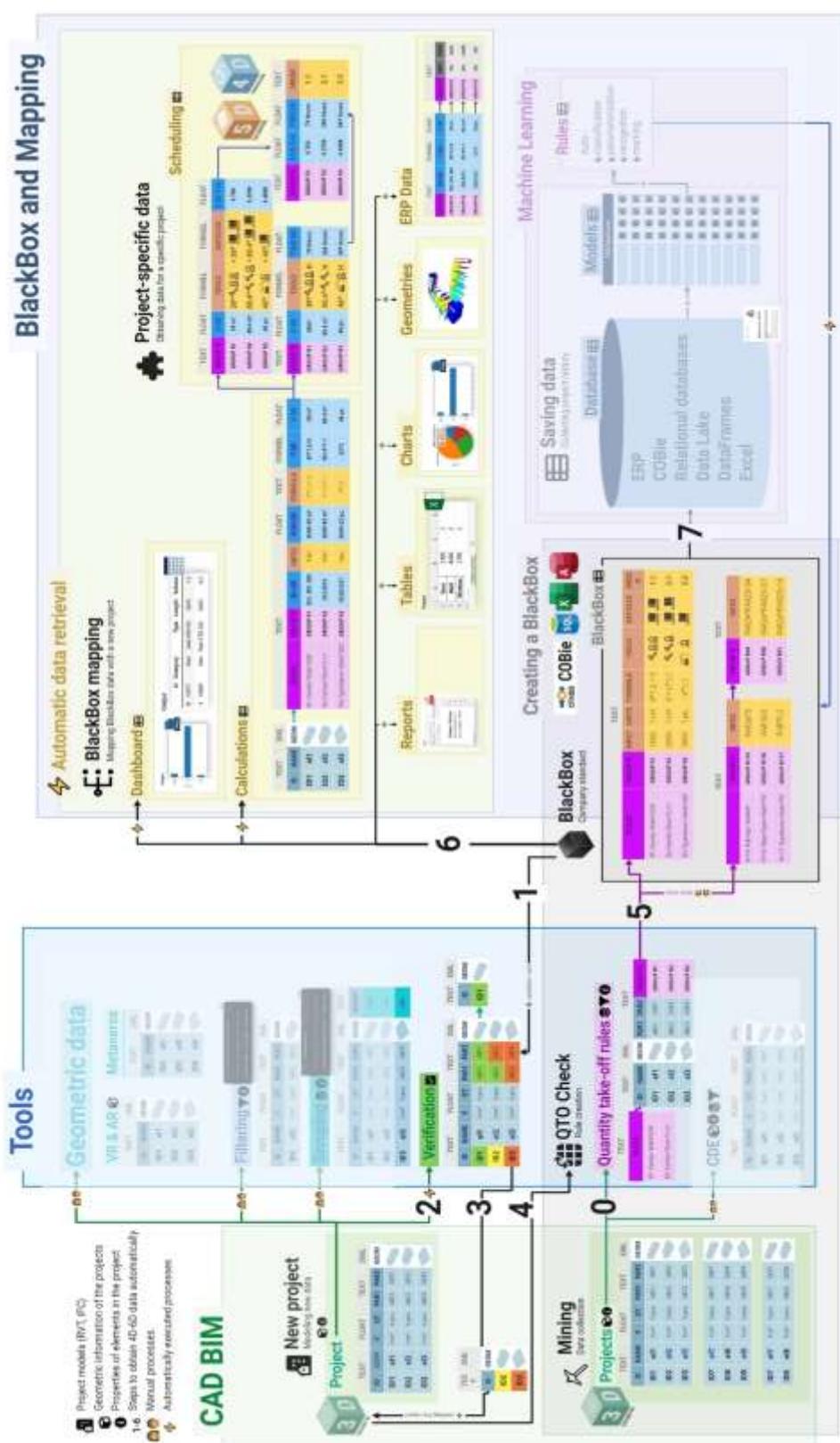


Фиг. 5.4-2 Архитектура на строителна ERP -система с CAD (BIM), от създаването на правила за валидиране (1) до автоматичното изчисляване на разходите и работните графици (5-6).

В една модулна ERP -система процесите се интегрират с помощта на софтуер, който включва потребителски интерфейс. Зад интерфейса се намира бек-енд, където структурирани таблици обработват данни, като извършват различни операции, предварително конфигурирани от мениджъра или администратора. В резултат на това потребителят, благодарение на предварително дефинираната и персонализирана логика за автоматизация (в модулите BlackBox /WhiteBox), получава полуавтоматично подгответи документи, които отговарят на неговите задачи.



Фиг. 5.4-3 ERP -системата помага на мениджърите и потребителите да се движат между специализираните таблици, за да генерират нови данни.



Фиг. 5.4-4 ERP -системата е интегрирана с аналитични инструменти и автоматизира процеса на вземане на решения в компанията.

По подобен начин процесите в ERP -системите, от създаването до окончателното изчисление (стъпки 1-6 Фигура 5.4-3), представляват верига от взаимосвързани стъпки, които в крайна сметка осигуряват прозрачност, ефективност и точност при планирането.

Съвременните строителни ERP -системи включват не само модули за изчисляване на разходите и графика, но и десетки други предварително конфигурирани модули, които обикновено обхващат управлението на документи, проследяването на напредъка на проекта, управлението на договори, веригата на доставки и логистиката, както и интеграцията с други бизнес системи и платформи. Интегрираните аналитични инструменти на ERP позволяват на потребителите да автоматизират създаването на информационни табла за наблюдение на ключовите показатели за ефективност на проекта (KPI - key performance indicators). По този начин се осигурява централизирано и последователно управление на всички аспекти на даден строителен проект, като се прави опит за обединяване на голям брой приложения и системи в една платформа.

В бъдеще ERP -аналитиката ще се използва в комбинация с машинно обучение, за да се увеличи точността и да се оптимизира процесът на изчисляване на бъдещите атрибути на проекта. Данныте и атрибутите, анализирани и събрани от ERP системите в Big Data (фиг. 5.4-4), в бъдеще ще формират основата за създаване на прогнозни модели, които могат точно да предвиждат потенциални закъснения, рискове или например възможни промени в разходите за материали.

Като алтернатива на ERP, в строителния бранш често се използва PMIS (Project Management Information System) - система за управление на проекти, предназначена за подробен контрол на задачите на ниво отделен строителен проект.

## PMIS: междуенно звено между ERP и строителната площадка

За разлика от ERP, която обхваща цялата верига от бизнес процеси на дадена компания, PMIS се фокусира върху управлението на конкретен проект, като следи за срокове, бюджети, ресурси и документация.

PMIS (Информационна система за управление на проекти) е софтуер за управление на строителни проекти, предназначен за планиране, проследяване, анализиране и отчитане на всички аспекти на проекта.

PMIS ви позволява да управлявате документи, графики, бюджети и На пръв поглед PMIS може да изглежда като дублиращо решение на ERP, но основната разлика е в нивото на управление:

- **ERP** е насочена към бизнес процесите на компанията като цяло: разходи, договори, поръчки, човешки ресурси и управление на ресурсите на корпоративно ниво.
- **PMIS** се фокусира върху управлението на отделни проекти, като осигурява подробно планиране, контрол на промените, отчитане и координация на участниците.

В много случаи ERP -системите вече разполагат с достатъчна функционалност и внедряването на PMIS става по-скоро въпрос на удобство и предпочтение на компанията. Много изпълнители и клиенти използват PMIS не защото това е необходимо, а защото е наложено от доставчика или от голям клиент, който иска да обобщава данни на определена платформа.

Трябва да се спомене, че в международната терминология за управление на строителни проекти

има и други отделни популярни понятия като PLM (Product Lifecycle Management) и EPC и EPC-M (Engineering, Procurement and Construction Management) - методи за договаряне в строителната индустрия.

Ако една компания вече използва ERP с модули за управление на проекти, въвеждането на PMIS може да се окаже ненужна връзка, която дублира функционалността. Ако обаче процесите не са автоматизирани и данните са фрагментирани, PMIS може да бъде по-удобен и лесен за поддръжка инструмент.

## Спекулации, печалба, изолираност и липса на прозрачност в ERP и PMIS

Въпреки външната простота на интерфейсите и процедурите, строителните ERP и PMIS -системи в повечето случаи са затворени и негъвкави решения. Такива системи обикновено се доставят като предварително конфигуриран софтуерен пакет от един доставчик, с ограничен достъп до вътрешните бази данни и логиката на процесите.

Доставчиците на CAD-(BIM-) все повече поемат разработването и контрола на такива системи, тъй като техните бази данни съдържат информацията, която се изисква от ERP системите: количествени и обемни характеристики на елементите на проекта. Вместо обаче да предоставят достъп до тези данни в отворен или машинночетим формат, доставчиците предлагат само ограничени потребителски сценарии и затворена логика на обработка - предварително дефинирана в рамките на модулите BlackBox. Това намалява гъвкавостта на системата и не позволява тя да бъде адаптирана към специфичните условия на проекта.

Ограниченната прозрачност на данните остава едно от основните предизвикателства пред цифровите процеси в строителството. Затворената архитектура на базите данни, липсата на достъп до пълните набори от атрибути на строителните елементи, фокусът върху модулите за автоматизация *BlackBox* и липсата на отворени интерфейси значително увеличават рисковете от бюрокрация на документите. Подобни ограничения създават пречки в процеса на вземане на решения, затрудняват проверката на информацията и отварят вратата за скриване на данни или спекулации в ERP/PMIS системите. Обикновено потребителите получават само ограничен достъп - било то орязан интерфейс или частичен API - без възможност за пряко взаимодействие с първичните източници на данни. Това е особено важно, когато става въпрос за параметри, генериирани автоматично от CAD проекти, като обеми, площи и количества, използвани за изчисления на QTO.

В резултат на това, вместо да търсят ефективност чрез автоматизация на процесите, отворени данни, намаляване на транзакционните разходи и създаване на нови бизнес модели, много строителни компании се фокусират върху управлението на външни параметри - манипулационни фактори, корекционни коефициенти и методи за изчисление, влияещи върху разходите по проекта в затворени ERP/PMIS платформи. Това създава възможност за спекулации, изкривява реалните производствени разходи и намалява доверието между всички участници в строителния процес.

В строителството печалбата се формира като разлика между приходите от завършен проект и променливите разходи, които включват проектиране, материали, труд и други преки разходи, пряко свързани с изпълнението на проекта. Ключовият фактор, който влияе върху стойността на тези разходи, обаче, не е само технологията или логистиката, но и бързината и точността на изчисленията, както и качеството на управлensките решения в компанията.

Проблемът се задълбочава от факта, че в повечето строителни компании процесите на изчисляване на разходите остават непрозрачни не само за клиентите, но и за самите служители, които не са част от отделите за оценка или финанси. Подобна затвореност благоприятства формирането в рамките на компанията на привилегирована група специалисти - носители на "финансова експертиза", които имат изключителното право да редактират атрибути и корекционни коефициенти в ERP/PMIS - системите. Тези служители, заедно с ръководителите на дружествата, могат реално да контролират финансата логика на проекта.

В такива условия оценителите се превръщат във "финансови жонгльори", които балансират между максимизирането на печалбата на компанията и необходимостта да поддържат конкурентна цена за клиента. В същото време те трябва да избягват явни и груби манипулации, за да не подкопаят репутацията на компанията. На този етап се определят коефициенти, за да се прикрият надценени обеми или разходи за материали и работи.

В резултат на това основната схема за повишаване на ефективността и рентабилността на компаниите, работещи в строителния бранш, не е автоматизацията и ускоряването на процесите на вземане на решения, а спекулациите с цените на материалите и строителните работи (фиг. 5.4-5). За-вишаването на стойността на работите и материалите се извършва чрез сиво счетоводство в затворените ERP /PMIS - системи, като се завишиват процентите над средните пазарни цени на материалите или обемите на работите с помощта на коефициенти (фиг. 5.1-6), които бяха разгледани в главата "Съставяне на калкулации и изчисляване на стойността на работите на базата на ресурсната база".

В резултат на това клиентът получава изчисление, което не отразява реалните разходи или обхвата на работата, а е производно на много скрити вътрешни коефициенти. В същото време подизпълнителите, в опит да отговорят на занижените цени, определени от главния изпълнител, често са принудени да купуват по-евтини и нискокачествени материали, което влошава крайното качество на строителството.

Спекулативният процес на търсене на печалби от въздуха в крайна сметка вреди както на клиентите, които получават ненадеждни данни, така и на изпълнителите, които са принудени да търсят все повече и повече спекулативни модели

В резултат на това, колкото по-голям е проектът, толкова по-високо е нивото на бюрокрация при управлението на данните и процесите. Всяка стъпка и всеки модул често крият непрозрачни коефициенти и надбавки, заложени в алгоритмите за изчисление и вътрешните процедури. Това не само

затруднява одита, но и значително изкривява финансовата картина на проекта. При големите строителни проекти подобни практики често водят до многократно (понякога до десетократно) увеличение на крайната цена, докато реалните обеми и разходи остават извън ефективния контрол на клиента (фиг. 2.1-3 Сравнение на планираните и реалните разходи на големи инфраструктурни проекти в Германия).

Според доклада на McKinsey & Company "Представяне на цифровото бъдеще на строителството" (2016 г.) големите строителни проекти се завършват средно с 20% по-късно от планираното и с до 80% над бюджета [107].

Отделите за прогнозиране и бюджетиране се превръщат в най-охраняваното звено в една компания. Достъпът до тях е строго ограничен дори за вътрешните специалисти, а поради затворената логика и структурите на базите данни е невъзможно да се оцени обективно и без изкривявания ефективността на проектните решения. Липсата на прозрачност води до това, че компаниите са принудени не да оптимизират процесите, а да се борят за оцеляване чрез "творческо" управление на цифрите и факторите (фиг. 5.3-1, фиг. 5.1-6 - например параметърът "Bid. Factor").



**Фиг. 5.4-5 Коефициентите на спекулация на ниво населено място са основната печалба на компаниите и изкуството да се жонгира между качеството на работата и репутацията.**

Всичко това поставя под съмнение продължаващото използване на затворени ERP/PMIS-системи в строителството. В контекста на цифровата трансформация и нарастващите изисквания за прозрачност от страна на клиентите (фиг. 10.2-3) е малко вероятно реализацията на проектите в дългосрочен план да остане зависима от затворени решения, които ограничават гъвкавостта, затрудняват интеграцията и пречат на развитието на бизнеса.

И колкото и да е благоприятно за строителните компании да работят със силози от данни и непрозрачни данни в затворени бази данни - неизбежно бъдещето на строителната индустрия ще включва преход към отворени платформи, машинно четивни и прозрачни структури от данни и автоматизация, основана на доверие. Тази трансформация ще бъде управлявана отгоре - под натиска на

клиентите, регуляторните органи и обществото, които все повече изискват отчетност, устойчивост, прозрачност и икономическа целесъобразност.

## Край на ерата на затворените ERP /PMIS: строителната индустрия се нуждае от нови подходи

Използването на обемисти модулни ERP/PMIS -системи, състоящи се от десетки милиони редове код, прави всякакви промени в тях изключително трудни. В този случай преходът към нова платформа при наличието на вече предварително конфигурирани за компанията модули, десетки хиляди статии в ресурсни бази данни (фиг. 5.1-3) и хиляди готови изчисления (фиг. 5.1-6) се превръща в скъп и продължителен процес. Колкото повече е кодът и наследената архитектура - толкова по-високо е нивото на вътрешна неефективност, а всеки нов проект само ще влошава нещата. В много компании миграцията на данни и интеграцията на нови решения се превръщат в многогодишни епопеи, съпроводени с постоянни преработки и безкрайно търсene на компромиси. Резултатът често е завръщане към стари, познати платформи, въпреки техните ограничения.

Както е подчертано в германския доклад Black Book [108] за системните провали в управлението на данните в строителството, фрагментарността на информацията и липсата на централизиран подход към нейното управление е основна причина за неефективност. Без стандартизация и интеграция данните губят своята стойност, превръщайки се в архив, а не в инструмент за управление.

Основна причина за загубата на качество на данните е неадекватното планиране и контрол на строителните проекти, което често води до значително увеличаване на разходите. В раздела "Фокус: Експлозия на разходите" на Черната книга се анализират основните фактори, които допринасят за тези нежелани последици. Те включват неадекватни анализи на нуждите, липса на предпроектни проучвания и некоординирано планиране, водещо до допълнителни разходи, които биха могли да бъдат избегнати.

В една зряла ИТ екосистема на компанията подмяната на останяла система е сравнима с подмяната на носеща колона във вече построена сграда. Не е достатъчно просто да се премахне старата и да се инсталира нова - важно е да се направи така, че сградата да остане стабилна, таваните да не се срутят и всички комуникации да продължат да работят. Именно тук се крие трудността: всяка грешка може да има сериозни последици за цялата система на компанията.

Въпреки това разработчиците на големи ERP продукти за строителната индустрия продължават да използват количеството написан код като аргумент в полза на своята платформа. На специализирани конференции все още могат да се чуят фрази от рода на: "Пресъздаването на такава система би отнело 150 човекогодини", въпреки че по-голямата част от функционалността на такива системи крие бази данни и доста прости функции за работа с таблици, опаковани в специален фиксиран потребителски интерфейс. На практика обемът на кода от "150 човекогодини" се превръща по-скоро в тежест, отколкото в конкурентно предимство. Колкото повече код - толкова по-високи са разходите за поддръжка, толкова по-трудно е адаптирането към новите условия и толкова по-висок е прагът за навлизане на нови разработчици и клиенти.

Много от модулните строителни системи днес приличат на тромави и оstarели "конструкции на Франкенщайн", при които всяка невнимателна промяна може да доведе до неуспех. Всеки нов модул увеличава сложността на и без това претоварената система, превръщайки я в лабиринт, в който могат да се ориентират само няколко специалисти, което допълнително затруднява поддръжката и модернизацията.

Сложността се осъзнава и от самите разработчици, които периодично правят паузи за рефакторинг - преразглеждане на архитектурата, за да се вземе предвид появата на нови технологии. Въпреки това, дори ако рефакторингът се извършва редовно, сложността неизбежно нараства. Архитектите на такива системи свикват с нарастващата сложност, но за новите потребители и специалисти тя се превръща в непреодолима бариера. В резултат на това целият експертен опит се концентрира в ръцете на няколко разработчици и системата престава да бъде мащабируема. В краткосрочен план тези експерти са полезни, но в дългосрочен план те се превръщат в част от проблема.

*Организациите ще продължат да интегрират "малки" данни с техните големи аналоги и е глупаво някой да вярва, че едно приложение - колкото и да е скъпо или надеждно - може да се справи с всичко [109].*

- Фил Саймън, водещ на подкаста Conversations About Collaboration

Възниква основателен въпрос: наистина ли се нуждаем от такива тромави и затворени системи за изчисляване на разходите и сроковете на работа под формата на таблици, ако други индустрии отдавна са в състояние да се справят с подобни задачи, използвайки аналитични инструменти с отворени данни и прозрачна логика?

В момента затворените модулни платформи все още са търсени в строителната индустрия, най-вече поради спецификата на отчитане на разходите (фиг. 5.1-7). Такива системи често се използват за изпълнение на сиви или непрозрачни схеми, което позволява реалните разходи да бъдат скрити от клиента. Въпреки това, тъй като индустрията узрява в цифрово отношение, предимно клиентите, и преминава в така наречената "ера на Uber", посредниците, а именно строителните компании с техните ERP, ще загубят значението си при изчисляването на времето и разходите. Това ще промени завинаги облика на строителната индустрия. Прочетете повече в последната част на книгата и в главата "Строителство 5.0: Как да печелим, когато вече не можем да се крием".

Хилядите наследени решения, натрупани през последните 30 години, с хиляди човекогодини, инвестиирани в разработката, ще започнат бързо да изчезват. Преходът към отворено, прозрачно и гъвкаво управление на данните е неизбежен. Единственият въпрос е кои компании ще успеят да се адаптират към тези промени и кои ще останат заложници на стария модел.

Подобна ситуация се наблюдава в областта на CAD (BIM) инструментите, чито данни днес попълват обемните параметри на проектантските единици в ERP/PMIS-системите. Първоначално идеята за BIM (разработена още през 2002 г. [110]) се основава на концепцията за единна интегрирана база данни, но на практика днес работата с BIM изисква цял набор от специализиран софтуер и формати.

Това, което се предполагаше, че ще опрости управлението на проектирането и строителството, се превърна в още един слой собственически решения, които усложняват интеграцията и намаляват гъвкавостта на бизнеса.

## Следващи стъпки: ефективно използване на данните от проекта

В тази част показваме как структурираните данни стават основа за точни изчисления на разходите и графика на строителните проекти. Автоматизирането на процесите на ОТО, планиране и оценка намалява разходите за труд и значително подобрява точността на резултатите.

За да обобщим тази част, си струва да подчертаем основните практически стъпки, които ще ви помогнат да приложите разгледаните подходи към ежедневните си задачи. Тези подходи са универсални - те са полезни както за цифровата трансформация на дадена компания, така и за ежедневната работа на специалистите, участващи в изчисленията:

### ■ Автоматизиране на рутинни изчисления

- Опитайте се да намерите стандартни калкулации на разходите за работа, които можете да използвате в работата си.
- Анализирайте кои методи се използват за остойностяване или изчисляване на разходите за строителни работи или процеси на строителната площадка във вашата страна (Фигура 5.1-7).
- Ако работите с CAD система - запознайте се с функциите за автоматично извлечане на спецификации и QTO данни във вашия CAD (BIM-) софтуер.
- Използвайте LLM, за да напишете чернова на код за автоматизиране на изчисленията

### ■ Разработване на собствени инструменти за QTO

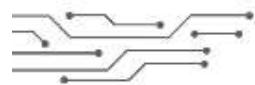
- Създаване на скриптове или таблици за автоматизиране на броенето на обеми
- Стандартизиране на категориите и групите елементи за последователен подход към оценката
- Документиране на методологията за изчисление, за да се осигури възпроизвежданост на резултатите при нови проекти.

### ■ Интегрирайте различни аспекти на проекта в работата си

- Ако работите с модулни системи, опитайте се да визуализирате процесите си не само като диаграми или графики, но и на ниво данни - особено под формата на таблици.
- Усвояване на автоматичното обединяване на данни, извлечени от бази данни на CAD, с изчисления - с код на Python, използващ групиране, филтриране и агрегиране
- Създаване на ясни визуализации на групи QTO за представяне на сложна информация на колеги и клиенти

Тези стъпки ще спомогнат за изграждането на устойчива система за изчисления, основана на автоматизация и стандартизация на данните. Този подход ще подобри точността и ще намали рутинните ежедневни проблеми при изчисленията.

Следващите глави са посветени на техническите аспекти на CAD - (BIM-) продуктите и причините, поради които CAD базите данни все още са трудни за интегриране в бизнес процесите на компаниите. Ако сега не се интересувате от историята на внедряването на BIM в строителството, еволюцията на CAD инструментите и техническите особености на работата с тези технологии, можете да преминете направо към седмата част на книгата "Вземане на решения, основани на данни".



## МАКСИМАЛНО УДОБСТВО С ПЕЧАТНАТА ВЕРСИЯ

Вие притежавате безплатна цифрова версия на **Data-Driven Construction**. За по-удобна работа и бърз достъп до материалите ви препоръчваме да обърнете внимание на **печатното издание**:



■ **Винаги под ръка:** книгата в печатен формат ще се превърне в надежден работен инструмент, който ще ви позволи бързо да намерите и използвате правилните визуализации и диаграми във всяка работна ситуация.

■ **Висококачествени илюстрации:** всички изображения и графики в печатното издание са представени с максимално качество.

■ **Бърз достъп до информация:** лесна навигация, възможност за правене на бележки, отметки и работа с книгата навсякъде.

Закупувайки пълната печатна версия на книгата, вие получавате удобен инструмент за комфортна и ефективна работа с информация: възможност за бързо използване на визуални материали в ежедневните задачи, бързо намиране на необходимите схеми и водене на бележки. Освен това с покупката си подкрепяте разпространението на отвореното знание.

Поръчайте печатна версия на книгата на адрес: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)





## VI ЧАСТ

# **CAD И BIM: МАРКЕТИНГ, РЕАЛНОСТ И БЪДЕЩЕТО НА ДАННИТЕ ЗА ПРОЕКТИРАНЕ В СТРОИТЕЛСТВОТО**

Шестата част на книгата представя критичен анализ на еволюцията на CAD и BIM -технологиите и тяхното въздействие върху процесите на управление на данни в строителството. Тя проследява историческата трансформация на концепцията BIM от първоначалната идея за интегрирана база данни до настоящите маркетингови конструкции, промотирани от производителите на софтуер. Оценява се влиянието на патентованите формати и затворените системи върху ефективността на обработката на данни за проектите и цялостното представяне на строителната индустрия. Подробно са анализирани проблемите със съвместимостта на различните CAD-системи и трудностите при интегрирането им с бизнес процесите на строителните компании. Обсъждат се настоящите тенденции към опростени отворени формати за данни, като например USD, и тяхното потенциално въздействие върху отрасъла. Представени са алтернативни подходи за извлечане на информация от затворени системи, включително техники за обратно инженерство. Анализирани са перспективите за прилагане на изкуствен интелект и машинно обучение за автоматизиране на процесите на проектиране и анализ на данни в строителството. Формулирани са прогнози за развитието на технологии за проектиране, ориентирани към реалните нужди на потребителите, а не към интересите на производителите на софтуер.

## ГЛАВА 6.1.

### **ПОЯВА НА КОНЦЕПЦИИТЕ НА BIM- В СТРОИТЕЛНАТА ИНДУСТРИЯ**

Първоначално тази шеста част, посветена на CAD (BIM), не беше включена в първата версия на книгата. Темите за патентованите формати, геометричните ядра и затворените системи са прекалено технически, претоварени с подробности и на пръв поглед безполезни за тези, които просто искат да разберат как да работят с данни. Отзовите иисканията за добавяне на разяснения към първата версия на книгата обаче показваха, че без разбиране на сложността на вътрешната работа на CAD системите, геометричните ядра, разнообразието от формати и несъвместимите схеми за съхранение на едни и същи данни е невъзможно да се разбере истински защо концепциите, насырчавани от производителите, често затрудняват работата с информацията и възпрепятстват прехода към отворено параметрично проектиране. Ето защо тази част е заета самостоятелно място в структурата на книгата. Ако CAD (BIM) не е приоритет за вас, можете да преминете направо към следващата част - "ЧАСТ VII: Вземане на решения, базирани на данни, анализи, автоматизация и машинно обучение".

#### **История на появата на BIM и отворен BIM като маркетингови концепции на доставчиците на CAD-**

С навлизането на цифровите данни през 90-те години на миналия век компютърните технологии бяха въведени не само в бизнес процесите, но и в процесите на проектиране, което доведе до появата на понятия като CAD (компютърно подпомагани системи за проектиране), а по-късно и BIM (информационно моделиране на сгради)

Както всяка иновация обаче, те не са крайната точка на развитието. Концепции като BIM се превърнаха във важен етап в историята на строителната индустрия, но рано или късно те могат да отстъпят място на по-добри инструменти и подходи, които ще отговорят по-добре на предизвикателствата на бъдещето.

Претоварена от влиянието на доставчиците на CAD и объркана от сложността на собственото си прилагане, концепцията за BIM, която се появи през 2002 г., може и да не доживее тридесетия си рожден ден, подобно на рок звезда, която изгрява ярко, но бързо угасва. Причината е проста: изискванията на специалистите по данни се променят по-бързо, отколкото доставчиците на CAD могат да се адаптират към тях.

Изправени пред липсата на качествени данни, днешните специалисти в строителната индустрия изискват междуплатформена оперативна съвместимост и достъп до отворени данни от проекти с CAD-, за да опростят техния анализ и обработка. Сложността на CAD данните и обърквашата им обработка оказват негативно влияние върху всички участници в строителния процес: проектанти, ръководители на проекти, строителни работници на обекта и в крайна сметка върху клиента.

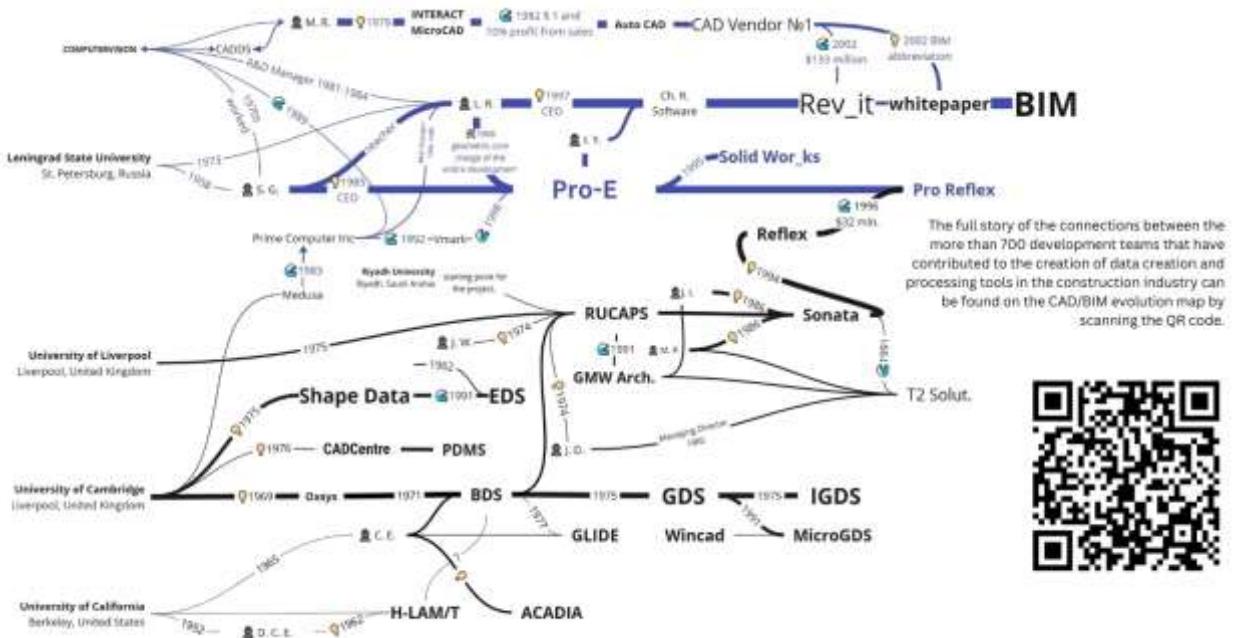
Вместо пълноценен набор от данни за експлоатация днес клиентът и инвеститорът получават контейнери в CAD- формати, които изискват сложни геометрични ядра, разбиране на схемите за данни,

ежегодно актуализирана API -документация и специализиран CAD софтуер (BIM) за работа с данните. В същото време голяма част от данните за проектиране остават неизползвани.

В днешния свят на проектиране и строителство сложността на достъпа до CAD данни води до свръхинженерно управление на проекта. Средните и големите компании, работещи с CAD данни или разработващи BIM -решения, са принудени или да поддържат тесни взаимоотношения с доставчиците на CAD решения за достъп до данните чрез API, или да заобикалят ограниченията на CAD доставчиците, като използват скъпи SDK конвертори за обратно инженерство, за да получат отворени данни [75].

Подходът, основан на собствени данни, е остатъл и вече не отговаря на изискванията на днешната цифрова среда. Бъдещето ще раздели компаниите на два вида: такива, които използват ефективно отворените данни, и такива, които напускат пазара.

Концепцията за BIM (Building Information Modeling), се появява в строителната индустрия с публикуването на един от основните доставчици на CAD - Whitepaper BIM [54] през 2002 г. и, допълнена от машиностроителната концепция BOM (Bills of Materials), произлиза от параметричния подход към създаването и обработката на данни за проекта (фиг. 6.1-1). Параметричният подход към създаването и обработката на проектни данни е един от първите, които се прилагат в системата Pro-E за машиностроително проектиране (MCAD). Тази система се превърна в прототип [111] много съвременни CAD решения, включително и за тези, които се използват днес в строителната индустрия.



Фиг. 6.1-1 Карта на историята на концепцията BIM и сходни концепции.

Журналистите и консултанти по AEC, които до началото на 2000 г. популяризираха CAD инструменти -доставчици, от 2002 г. насочиха вниманието си към Whitepaper BIM. Именно Whitepaper BIM 2002-2004 и статиите, публикувани през 2002 г., 2003 г., 2005 г. и 2007 г., изиграха ключова роля за популяризирането на концепцията BIM в строителната индустрия [112].

*Информационното моделиране на сгради е стратегия..... [име на фирмата доставчик на CAD] за прилагане на информационните технологии в строителната индустрия.*

- BIM Whitepaper, 2002 [60]

Към средата на 2000-те години "изследователите" започват да свързват концепцията BIM-, публикувана от доставчика на CAD- през 2002 г., с по-ранни научни трудове, като например BDS на Чарлз Ийстман, които стават основа за системи като GLIDE, GBM, BPM, RUCAPS. В своя пионерски труд Building Description System (1974 г.) Чарлз Истман поставя теоретичните основи на съвременното информационно моделиране. Терминът "база данни " се появява 43 пъти в неговия труд (фиг. 6.1-2) - по-често от всеки друг, с изключение на думата "сграда".

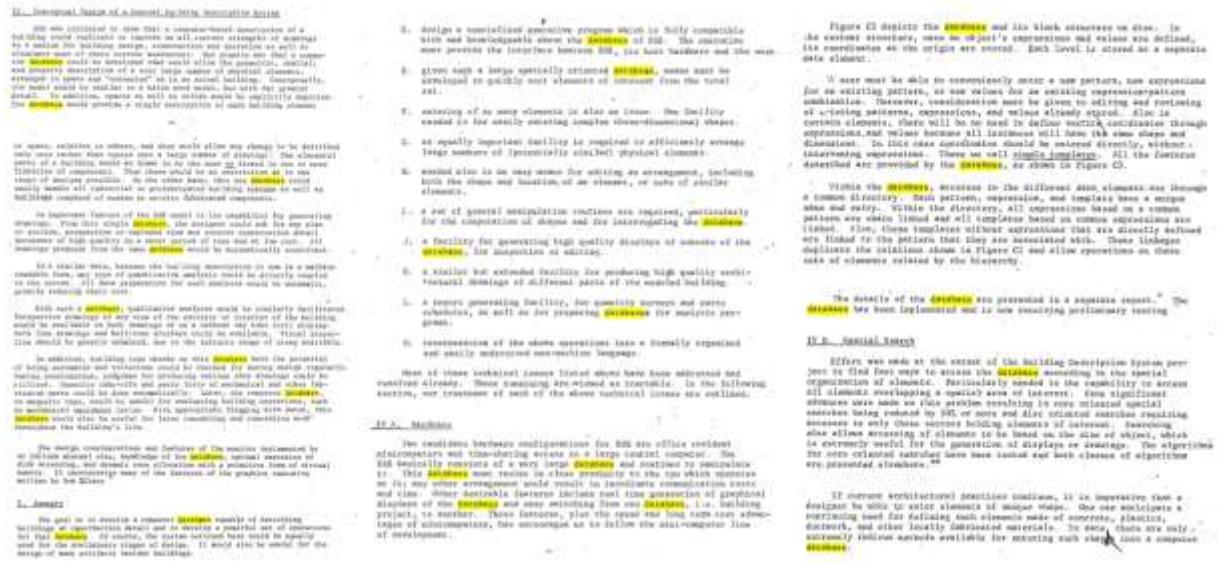
Ключовата идея на Ийстман е, че цялата информация за сградата - от геометрията до свойствата на елементите и техните взаимовръзки - трябва да се съхранява в една структурирана база данни. Именно от тази база данни могат автоматично да се генерират и анализират чертежи, спецификации, изчисления и съответствие с правилата.

Eastman изрично критикува чертежите като остатъл и излишен метод за комуникация, като посочва дублирането на информация, проблемите с актуализирането и необходимостта от ръчни актуализации при извършване на промени. Вместо това той предложи единен цифров модел в база данни, където всяка промяна се прави веднъж и автоматично се отразява във всички изгледи.

Заслужава да се отбележи, че в своята концепция Ийстман не поставя визуализацията в центъра. Централно място в неговата система заемаше информацията: параметри, връзки, атрибути, възможности за анализ и автоматизация. Чертежите в неговото разбиране са били само една от формите за показване на данните от базата данни, а не основният източник на информация за проектирането.

В първата бяла книга за BIM на водещия доставчик на CAD изразът "база данни " се използва толкова често, колкото и в БДС на Чарлз Ийстман - 23 пъти [60] седем страници и е една от най-популярните думи в документа след "сграда", "информация", "моделиране" и "проектиране". Към 2003 г. обаче терминът "база данни" се появява само два пъти в подобни документи [61], а към края на 2000 г. темата за базите данни на практика изчезва от обсъждането на данните за проектиране. В резултат на това концепцията за "единна интегрирана база данни за визуален и количествен анализ" така и не се реализира напълно.

По този начин строителната индустрия премина от прогресивната концепция за БДС на Чарлз Ийстман с нейния акцент върху базите данни и идеите на Самуел Гайсберг за автоматично актуализиране на проектните данни от базите данни в машиностроителния продукт Pro-E (предшественик на популярните CAD -решения, използвани днес в строителството) до актуалния на пазара BIM, където управлението на данни чрез бази данни почти не се споменава, въпреки че това е концепцията, стояща зад първоначалния теоретичен



**Фигура 6.1-2 В концепцията за БДС, описана от Чарлз Иестман през 1974 г., фразата "база данни" (подчертана в жълто) е използвана 43 пъти.**

BDS и подобни концепции до 2000 г. са разработени по-скоро като цифрова база данни на сгради, отколкото като инструмент за визуализация. През 2002 г. BIM се превърна в инструмент за проектиране, при който базата данни остана на заден план. Какво загубихме при прехода от БДС и подобни концепции през 90-те години на миналия век към БИМ до средата на 2010 г.:

- Отворени бази данни: BDS и други подобни концепции наблюгат на анализа, BIM наблюга на проектирането.
- Гъвкавост при работа с данни: BDS наблюга на анализа на данни, BIM наблюга на процесите, които трябва да се основават на неясни данни.
- Прозрачност: BDS е замислен като отворена интегрирана база данни, докато доставчиците на CAD в BIM са направили своите бази данни напълно затворени и вече 20 години безуспешно се борят срещу инструментите за обратно инженерство, които отварят патентовани формати.

През последните 30 години проектантите никога не са имали достъп до "интегрирана база данни" и след двадесет години маркетингова еуфория около BIM-инструментите строителната индустрия започва да осъзнава последствията от тази мода.

## Реалността на BIM: вместо интегрирани бази данни - затворени модулни системи

Вместо да се съсредоточат върху данните, да ги структурират и интегрират в единни процеси, потребителите на CAD (BIM-) системи са принудени да работят с фрагментиран набор от патентовани решения, всяко от които диктува свои собствени правила на играта:

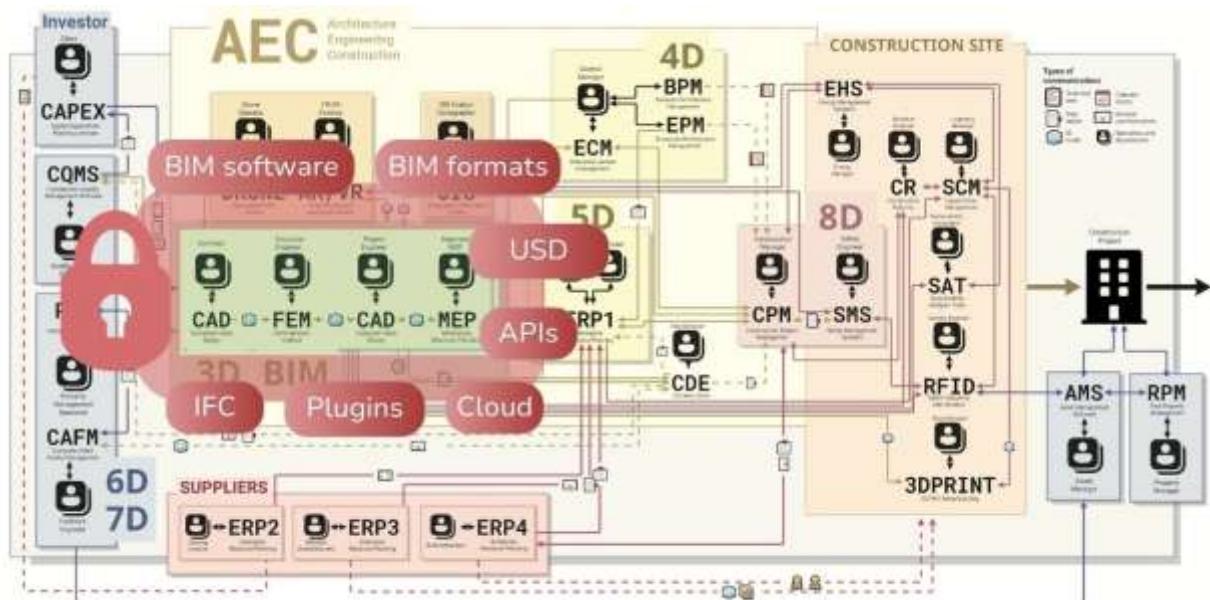
- **Единната база данни**, която беше предмет на първата Бяла книга за BIM, си остана мит. Въпреки гръмките твърдения, достъпът до данните все още е ограничен и разпределен между затворени системи.

- **BIM - моделите са се превърнали в затворена екосистема**, а не в инструмент. Вместо прозрачен обмен на информация, потребителите са принудени да плащат за абонаменти и да използват патентовани API.
- **Данните принадлежат на доставчиците, а не на потребителите.** Информацията за проектите е заключена в патентовани формати или облачни услуги, а не е достъпна в отворени и независими формати.

Инженерите-проектанти и ръководителите на проекти често нямат достъп нито до CAD базите данни, нито до формата, в който се съхраняват техните собствени данни за проекта. Това прави невъзможна бързата проверка на информацията или формулирането на изисквания към структурата и качеството на данните (фиг. 6.1-3). Достъпът до такива данни изисква цял набор от специализирани програми, свързани чрез API и плъгини, което води до прекомерна бюрократизация на процесите в строителната индустрия. В същото време тези данни се използват едновременно от десетки информационни системи и стотици специалисти.

*Трябва да можем да управляем всички тези данни [CAD (BIM)], да ги съхраняваме в цифров вид и да продаваме софтуер за управление на жизнения цикъл и процесите, защото за **всеки инженер** [проектант], който създава нещо [в CAD софтуер], **има десет души, които работят с тези данни**" [41].*

- Изпълнителен директор на CAD - доставчикът, създал концепцията за BIM, 2005 г.



Фиг. 6.1-3 CAD- (BIM-) базите данни остават една от последните затворени системи за ИТ отделите и мениджърите на данни в екосистемата на строителния бизнес.

Когато стане ясно, че BIM е по-скоро средство за комерсиализиране на бази данни, отколкото пълноценен инструмент за управление на бази данни, възниква логичният въпрос: как да се възстанови контролът върху данните? Отговорът е да се използват отворени структури от данни, при които потребителят, а не доставчикът на софтуера, става собственик на информацията.

Потребителите и разработчиците на решения в строителния бранш, подобно на колегите си от други отрасли, неизбежно ще се откажат от неясната терминология на доставчиците на софтуер, която доминираше през последните 30 години, и ще се съсредоточат върху ключовите аспекти на цифровизацията - "данни" и "процеси".

В края на 80-те години на миналия век ключовата област на цифровото развитие в строителството беше представена като въпрос на достъп до данни и управление на информацията за проекта. С течение на времето обаче фокусът се измести. Вместо да се разработват прозрачни и достъпни подходи за работа с данни, форматът IFC и концепцията за отворен BIM бяха активно популяризиращи като опити да се отклони вниманието на специалистите от темите за управление на проектните бази данни.

## **Появата на отворения формат IFC в строителната индустрия**

Така нареченият отворен формат IFC (Industry Foundation Classes) е позициониран като стандарт за осигуряване на оперативна съвместимост между различни CAD (BIM-) системи. Разработването му е извършено в рамките на организации, които са създадени и контролирани от основните производители на CAD. Въз основа на формата IFC през 2012 г. две CAD- компании разработиха маркетинговата концепция OPEN BIM[63].

**IFC** (Industry Foundation Classes) е отворен стандарт за обмен на данни в строителната индустрия, пред назначен да осигури оперативна съвместимост между различни CAD - (BIM-) системи.

Концепцията за отворен BIM - включва работа с информация от CAD бази данни и обмен на информация между системите чрез отворен формат за обмен на CAD данни - IFC.

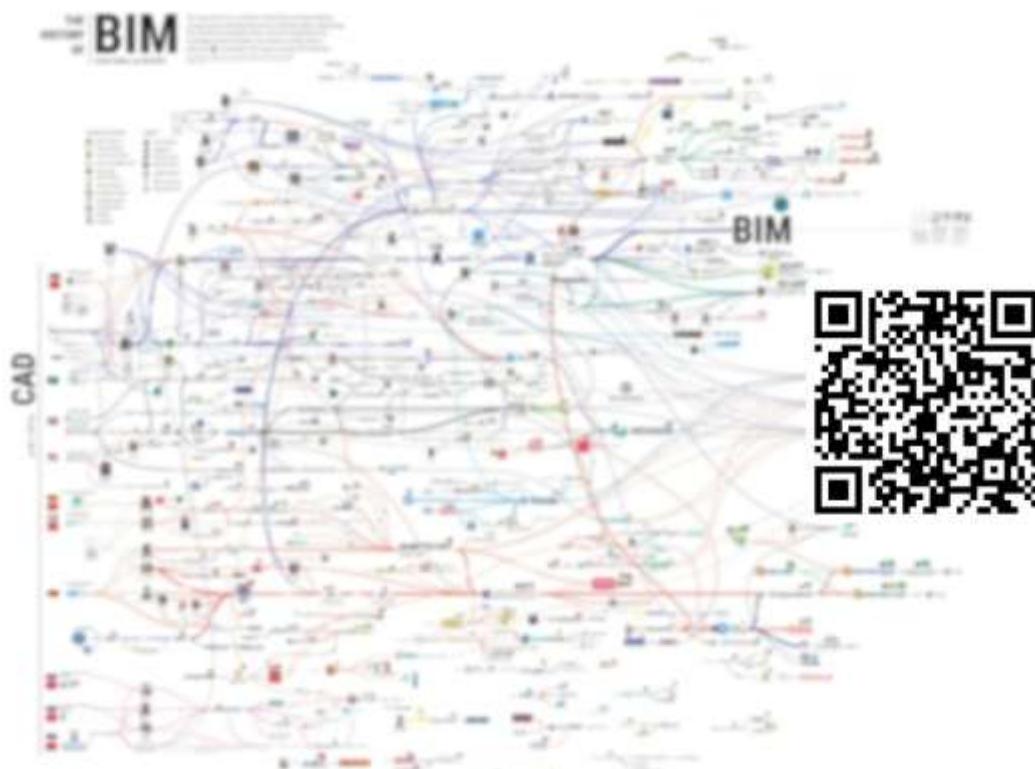
Програмата "Open BIM" е маркетингова кампания, инициирана от... [1 CAD доставчик],... [2 CAD доставчик] и други компании, за да се настърчи и улесни глобалното координирано популяризиране на концепцията OPEN BIM в цялата индустрия на AEC, с последователна комуникация и обща марка, достъпна за участниците в програмата.

- От уеб сайта на доставчика на CAD, програма OPEN BIM, 2012 г. [113]

IFC е адаптиран от Техническия университет в Мюнхен от машиностроителния формат STEP в края

на 80-те години на миналия век, а по-късно е регистриран от голяма проектантска компания и голям доставчик на CAD-, за да се създаде IAI (Industry Alliance for Interoperability) през 1994 г. [114] (фиг. 6.1-4). Форматът IFC е разработен, за да осигури оперативна съвместимост между различни CAD системи и се основава на принципите, заложени в машиностроителния формат STEP, който от своя страна възниква от формата IGES, създаден през 1979 г. от група потребители и доставчици на CAD с подкрепата на NIST (The National Institute of Standards and Technology) и Министерството на отбраната на САЩ [115].

Въпреки това сложната структура на IFC, тясната му зависимост от геометричното ядро, както и несъответствията при прилагането на формата от различните софтуерни решения доведоха до много проблеми при практическото му прилагане. Подобни трудности - загуба на детайли, ограничаване на точността и необходимост от използване на междинни формати - са срещани преди това от специалистите по машиностроеене при работа с форматите IGES и STEP, от които е произлязъл IFC.



Фиг. 6.1-4 Карта на връзките между развойните екипи и CAD продуктите (BIM) [116].

През 2000 г. същият CAD -доставчик, който регистрира формата IFC и създава организацията IAI (по-късно bS), публикува бялата книга "Integrated Design and Manufacturing: Benefits and Rationale". [65]. В документа се подчертава значението на запазването на пълната детайлност на данните при обмена между програмите в рамките на една и съща система, без да се използват неутрални формати като IGES, STEP [идентични с IFC]. Вместо това беше предложено приложението да имат пряк достъп до основната CAD база данни, за да се предотврати загубата на точност на информацията.

През 2002 г. същият доставчик на CAD купува параметричен продукт BOM (фиг. 3.1-18, повече подробности в от третата част) и на негова основа формира концепцията BIM. В резултат на това при обмена на данни за строителни проекти се използват само затворени CAD формати или IFC формат (STEP), за чиито ограничения пише самият CAD доставчик през 2000 г., който въвежда този формат в строителната индустрия.

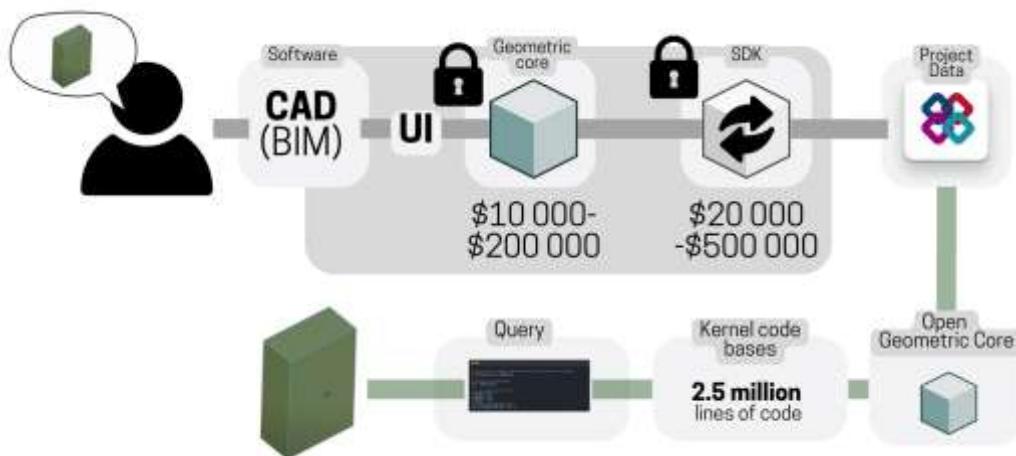
Подробна история на взаимодействието на повече от 700 развойни екипа, участващи в създаването на инструменти за създаване и обработка на строителни данни, е представена в картата "Еволюция на CAD (BIM)" [116]. [116].

Отворената форма на IFC се състои от геометрично описание на елементите на дизайна и описание на метаинформацията. Използват се различни методи за представяне на геометрията във формат IFC, като CSG и Swept Solids: въпреки това параметричното представяне BREP се е превърнало във водещ стандарт за прехвърляне на геометрията на елементите във формат IFC, тъй като този формат се поддържа при експортиране от CAD- (BIM-) програми и позволява евентуално редактиране на елементите при импортиране на IFC обратно в CAD програми.

## **Проблем с формата IFC в зависимост от геометричното ядро**

В повечето случаи, когато геометрията в IFC е дефинирана параметрично (BREP), става невъзможно визуализирането или извлечането на геометрични свойства, като например обем или площ на проектните обекти, само с помощта на IFC файл, тъй като за работа и визуализиране на геометрията в този случай е необходимо геометрично ядро (фигура 6.1-5), което първоначално липсва.

**Геометричното ядро** е софтуерен компонент, който осигурява основни алгоритми за създаване, редактиране и анализиране на геометрични обекти в CAD (CAD), BIM и други инженерни приложения. То отговаря за изграждането на двуизмерна и триизмерна - геометрия и за операции върху нея, като например: булеви операции, изглаждане, пре-сичания, трансформации и визуализация.

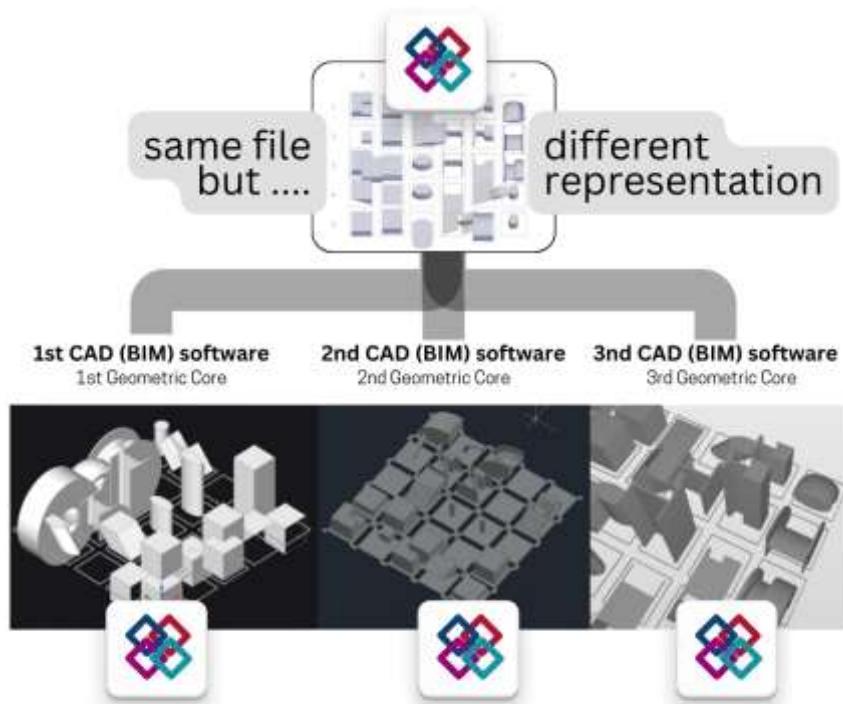


**Фигура 6.1-5 Създаването на геометрия чрез CAD- софтуер днес става чрез собственически геометрични ядра и SDK, които често не са собственост на доставчиците на CAD.**

Всяка CAD програма и всички програми, работещи с параметрични или IFC формати, имат собствено или закупено геометрично ядро. И ако при примитивните елементи във формат IFC -BREP не може да има проблеми и в програми с различни геометрични ядра тези елементи могат да се изобразяват по подобен начин, но освен проблемите с различните двигатели на геометричните ядра, има достатъчно елементи, които имат свои особености за правилно изобразяване. Този проблем е разгледан подробно в международното изследване "A reference study of IFC software support", публикувано 2019 г. [117].

*Едни и същи стандартизиран набори от данни дават противоречиви резултати, като са открити само няколко общи модела, а при поддържането на стандарта са установени сериозни проблеми [IFC], които вероятно се дължат на много големата сложност на стандартния модел на данните. Тук отчасти са виновни самите стандарти, тъй като те често оставят някои детайли неопределени, с висока степен на свобода и различни възможни тълкувания. Те допускат висока сложност в организацията и съхранението на обектите, което не благоприятства ефективното универсално разбиране, уникалните реализации и последователното моделиране на данни [117].*

- Референтно проучване на софтуерната поддръжка на IFC, 2021 г.



Source: Reference study of IFC software support: the GeoBIM benchmark 2019 — Part I <https://arxiv.org/pdf/2007.10951>

**Фиг. 6.1-6 Различните геометрични ядра дават различни представления на една и съща геометрия, описана параметрично (по [117]).**

Правилното разбиране на "някои разпоредби" е достъпно за платените членове на специалните организации, които разработват МФК. Вследствие на това, който иска да получи достъп до важни знания за определени функции на IFC, ще се опита да си сътрудничи с големите доставчици на CAD-или да достигне до качествено разглеждане на функциите чрез собствени проучвания

Натъквате се на въпрос за импортиране и експортиране на данни чрез формата IFC и питате колегите си: "Защо в IFC файла има информация за параметрично прехвърляне на помещения? В отворената спецификация не се казва нищо за това". Отговор от "по-знаещите" европейски доставчици: "Да, не е казано, но е позволено".

- От интервюто на разработчика на CAD 2021 [118]

IFC описва геометрията чрез параметрични примитиви, но не съдържа вградено ядро - неговата роля се изпълнява от CAD програмата, която компилира геометрията чрез геометричното ядро. Ядрото на геометрията извършва математическите изчисления и определя пресечните точки, докато IFC предоставя само данните за неговата интерпретация. Ако IFC съдържа некоректни лица, различните програми с различни геометрични ядра могат или да ги игнорират, или да генерират грешки в зависимост от ядрото.

В резултат на това, за да се работи с формат IFC, е необходимо да се отговори на основния въпрос, на който е трудно да се намери еднозначен отговор - какъв инструмент, с какво геометрично ядро трябва да се използва, за да се получи качеството на данните, което проектът първоначално е имал в CAD програмата, от която е получен IFC?

Проблемите с качеството на данните и сложността на формата IFC не позволяват директното използване на данните за проектиране за автоматизация на процесите, анализ и обработка на данни, което често води разработчиците до неизбежната необходимост да използват затворени CAD-решения с "качествен" достъп до данните [63], за което пише самият доставчик, регистрирал IFC през 1994 г. [65].

Всички особености на картографирането и генерирането на параметрите на IFC в геометричното ядро могат да бъдат реализирани само от големи екипи от разработчици, които имат опит в работата с геометрични ядра. Ето защо настоящата практика на особеностите и сложността на IFC формата е изгодна предимно за производителите на CAD-и и има много общо със стратегията на големите производители на софтуер "приеми, разшири, унищожи", когато нарастващата сложност на стандарта всъщност създава бариери за малките участници на пазара [94].

Стратегията на големите доставчици при подобна стратегия може да бъде да адаптират отворените стандарти, да добавят собствени разширения и функции, за да създадат зависимост на потребителите от своите продукти и след това да изтласкат конкурентите.

Форматът IFC, замислен като универсален мост между различни CAD- (BIM-) системи, в действителност изпълнява ролята на индикатор за проблеми със съвместимостта между геометричните ядра на различни CAD платформи, подобно на формата STEP, от който първоначално е произлязъл.

В резултат на това днес пълното и висококачествено внедряване на онтологията IFC е осъществимо за големите доставчици на CAD, които могат да инвестират значителни ресурси за поддръжка на всички същности и тяхното съпоставяне със собственото им вътрешно геометрично ядро, което не съществува за IFC като стандарт. Големите доставчици също така имат възможност да координират помежду си технически подробности за функции, които може да не са достъпни дори за най-активния участник в организациите за разработване на формата IFC.

За малките независими екипи и проектите с отворен код, които се стремят да подкрепят разработването на оперативно съвместими формати, липсата на собствено геометрично ядро се превръща в сериозен проблем. Без него е практически невъзможно да се вземат предвид всички различни тънкости и нюанси, свързани с междуплатформенния обмен на данни.

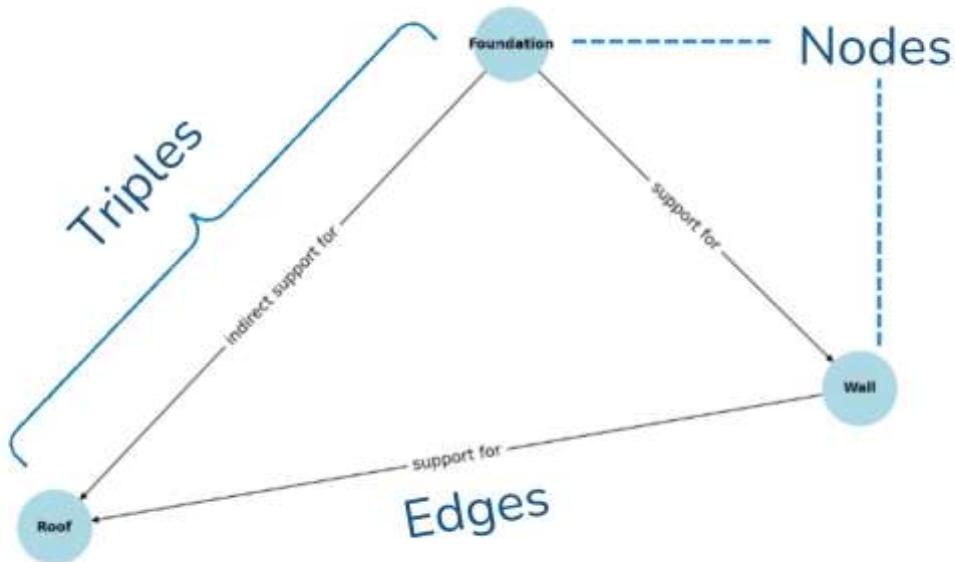
С разработването на параметричния формат IFC и отворената концепция за BIM, в строителния бранш се засилиха дискусиите за ролята на онтологията и семантиката в управлението на данните и процесите.

## Появяване в конструкцията на темата за семантичката и онтологията

Благодарение на идеите на семантичния интернет края на 90-те години на миналия век и усилията на организацията, участващи в разработването на формата IFC, семантичката и онтологиите се превърнаха в един от ключовите елементи на стандартизацията, обсъждана в строителната индустрия до средата на 20-те години.

Семантичните технологии са обединяване, стандартизиране и модифициране на големи масиви от разнородни данни и осъществяване на сложни търсения.

За съхраняване на семантични данни се използва OWL (Web Ontology Language), представен като RDF графи-триплети (Resource Description Framework) (фиг. 6.1-7). OWL се отнася до графови модели на данни, чиито видове разглеждахме по-подробно в главата "Модели на данни: връзки между данните и връзки между елементите".



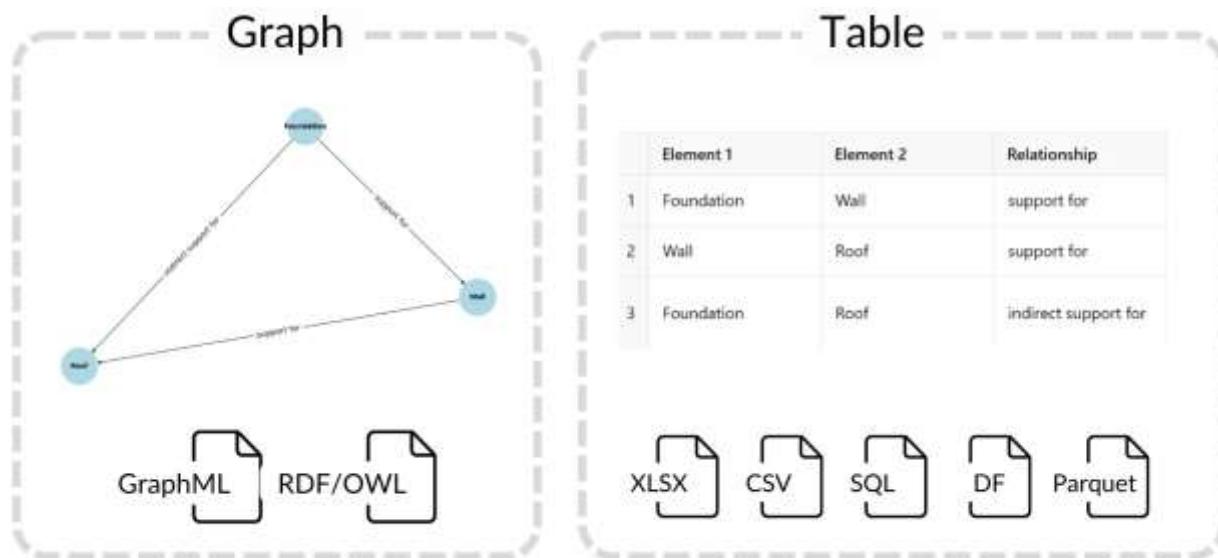
Фиг. 6.1-7 RDF модел на данни: възли, ребра и тройки, илюстриращи връзките между градивните елементи.

Теоретично, логическият извод на ризонери (софтуер за автоматичен логически извод) позволява да се извличат нови твърдения от онтологии. Например, ако в строителната онтология е записано, че "фундамент е опора за стена" и "стена е опора за покрив" (фиг. 6.1-7), ризонирацият модул е в състояние автоматично да заключи, че "фундамент е опора за покрив".

Такъв механизъм е полезен за оптимизиране на анализа на данни, тъй като избягва изричното предписване на всички зависимости. Въпреки това той не създава нови знания, а само идентифицира и структурира вече известни факти.

Семантиката сама по себе си не създава нов смисъл или знание и не превъзхожда другите технологии за съхранение и обработка на данни в този аспект. Представянето на данни от релационни бази данни като триплети не ги прави по-значими. Замяната на таблиците с графови структури може да е полезна за унифициране на моделите на данните, лесно извлечане и сигурно редактиране, но не прави данните "по-умни" - компютърът не започва да разбира по-добре съдържанието им.

Логическите връзки в данните могат да се организират без сложни семантични технологии (фиг. 6.1-8). Традиционните релационни бази данни (SQL), както и форматите CSV или XLSX позволяват изграждането на подобни зависимости. Например в колонарна база данни можете да добавите поле "поддръжка на покрива" и автоматично да свържете покрива с основата при създаването на стена. Този подход е реализиран без използването на RDF, OWL, графики или ризонери, като остава просто и ефективно решение за съхраняване и анализ на данни.



Фиг. 6.1-8 Сравнение на моделите на данни с графики и таблици за представяне на едни и същи логически връзки.

Решението на редица големи строителни компании и на организацията за разработване на формата IFC [94] да следват концепцията за семантична мрежа, която изглеждаше обещаваща в края на 90-те години на миналия век, оказа значително влияние върху разработването на стандарти в строителната индустрия.

Парадоксът обаче е, че самата концепция за семантична мрежа, първоначално предназначена за интернет, не е широко възприета дори в родната си среда. Въпреки развитието на RDF и OWL, пълноценната семантична мрежа не се е появила в първоначалния си замисъл, а създаването ѝ вече е малко вероятно.

## Защо семантичните технологии не оправдават очакванията в строителната индустрия

Други отрасли се сблъскват с ограниченията на технологиите за използване на семантиката. В игралната индустрия опитите да се опишат игровите обекти и техните взаимодействия чрез онтологии се оказаха неефективни поради високата динамика на промените. В резултат на това бяха предпочетени по-прости формати за данни като XML и JSON, заедно с алгоритмични решения. Подобна е ситуацията и в сектора на недвижимите имоти: поради регионалните различия в терминологията и честите промени на пазара използването на онтологии се оказва прекалено сложно, докато простите бази данни и стандарти като RETS [119] са в състояние да се справят по-добре с предизвикателствата на обмена на данни.

Техническите трудности, като например сложността на маркирането, голямата трудоемкост на поддръжката и ниската мотивация на разработчиците, забавиха приемането на семантичната мрежа и в други сектори на икономиката. RDF (Resource Description Framework) не се превърна в масов стандарт, а онтологиите се оказаха твърде сложни и икономически неоправдани.

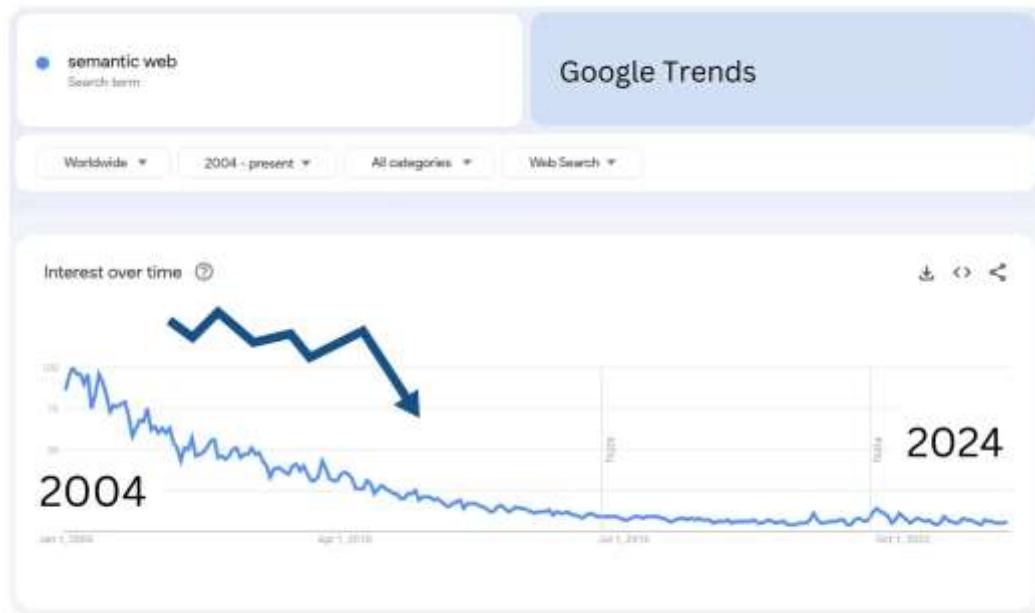
В резултат на това амбициозната идея за създаване на глобална семантична мрежа не успя да се осъществи. Въпреки че някои елементи на технологията, като например онтологиите и SPARQL, намериха своето място в корпоративните решения, първоначалната цел за създаване на единна цялостна структура от данни не беше постигната.

Концепцията за интернет, в който компютрите са в състояние да осмислят съдържанието, се оказа технически трудна и търговски неизгодна. Ето защо компаниите, които подкрепиха идеята, в крайна сметка ограничиха използването ѝ до отделни полезни инструменти, като оставиха RDF и OWL за тясно специализирани корпоративни нужди, а не за интернет като цяло. Анализът на тенденциите в Google (фигура 6.1-9) през последните 20 години показва, че може би няма повече перспективи за семантичната мрежа.

*Не е необходимо да умножавате ненужно единиците. Ако има няколко логически последователни обяснения на дадено явление, които го обясняват еднакво добре, при равни други условия трябва да се предпочете най-простото от тях.*

- Бръсначът на

Тук възниква логичен въпрос: защо изобщо да се използват триплети, ризери и SPARQL в конструкцията, когато можете да обработвате данни с помощта на популярни структурирани заявки (SQL, Pandas, Apache®)? В корпоративните приложения SQL е стандартът за работа с бази данни. SPARQL, напротив, изисква сложни графови структури и специализиран софтуер и според тенденциите в Google не привлича интереса на разработчиците.



**Фиг. 6.1-9** Интерес към заявките за "семантичен интернет" според статистиката на Google.

Бази данни с графики и класификационни дървета могат да бъдат полезни в някои случаи, но прилагането им не винаги е оправдано за повечето ежедневни задачи. В резултат на това създаването на графи на знанието и използването на технологиите на семантичния уеб има смисъл само когато е необходимо да се обединят данни от различни източници или да се реализират сложни логически заключения.

Преминаването от таблици към графични модели на данни подобрява търсенето и унифицира потока от информация, но не прави данните по-значими за машините. Въпреки че не е дали семантичните технологии трябва да се използват, а къде те наистина имат значение. Преди да внедрите онтологията, семантиката и графичните бази данни във вашата компания, разберете кои компании вече успешно използват тези технологии и къде са се провалили.

Въпреки амбициозните очаквания, семантичните технологии така и не се превърнаха в универсално решение за структуриране на данни в строителната индустрия. На практика тези технологии не доведоха до универсално решение, а само добавиха нови сложности и тези усилия повторят нереализираните амбиции на концепцията за семантичен интернет, където очакванията далеч надхвърлиха реалността.

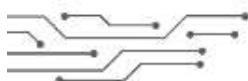


Фиг. 6.1-10 Геометрия и информация в строителните процеси: от сложни CAD и BIM- системи до опростени данни за анализ.

Докато в областта на информационните технологии неуспехите на семантичната мрежа бяха компенсирани от появата на нови технологии (големи данни, интернет на нещата, машинно обучение, AR/VR), строителната индустрия няма такива поводи.

В допълнение към предизвикателствата, свързани с използването на понятия за съобщаване на връзките на данните между елементите на проекта, остава и един основен проблем - самата наличност на тези данни. В строителния бранш все още преобладават затворените системи, което затруднява работата с данни, обмена на информация и подобряването на ефективността на процесите.

Именно затвореният характер на данните се превръща в една от основните пречки, които възпрепятстват разработването на цифрови решения в строителството. За разлика от ИТ индустрията, където отворените и хармонизираните формати за данни са се превърнали в стандарт, в сектора на CAD (BIM) всеки софтуер използва свой собствен формат, което създава затворени екосистеми и изкуствено ограничава потребителите.



## ГЛАВА 6.2.

# ЗАТВОРЕНИ ФОРМАТИ НА ПРОЕКТИТЕ И ПРОБЛЕМИ С ОПЕРАТИВНАТА СЪВМЕСТИМОСТ

## Затворени данни и спадаща производителност: задънената улица на CAD индустрията (BIM)

Собственият характер на CAD -системите е довел до факта, че всяка програма има свой собствен уникален формат на данните, който е или затворен и недостъпен отвън - RVT, PLN, DWG, NDW, NWD, SKP, или е достъпен в полуструктурен вид чрез доста сложен процес на преобразуване - JSON, XML (CPIXML), IFC, STEP и IfcXML, IfcJSON, BIMJSON, IfcSQL, CSV и т.н...

Различните формати на данни, в които могат да се съхраняват едни и същи данни за едни и същи проекти, се различават не само по структура, но и включват различни версии на вътрешните маркировки, които разработчиците трябва да вземат предвид, за да осигурят съвместимост на приложението. Например, CAD формат от 2025 г. ще се отвори в CAD програма от 2026 г., но същият проект никога няма да се отвори във всички версии на CAD програмата, които може да са били налични преди 2025 г.

Като не предоставя директен достъп до базите данни, доставчикът на софтуер в строителния бранш често създава свой собствен уникален формат и инструменти, които специалистът (инженер-проектант или мениджър на данни) трябва да използва за достъп, импортиране и експортиране на данни.

Вследствие на това доставчиците на базови CAD (BIM) и свързани с тях решения (например ERP/PMIS)) непрекъснато повишават цените за използване на продуктите, а обикновените потребители са принудени да плащат "комисиона" на всеки етап от трансфера на данни по формати [63]: за свързване, импортиране, експортиране и работа с данни, които потребителите сами са създали.

Разходите за достъп до данни в облачно хранилище от популярни CAD - (BIM-) продукти ще достигнат 1 доллар на транзакция през 2025 г. [120], а абонаментите за строителни ERP -продукти за средни компании достигат пет- и шестцифрени суми годишно [121].

Същността на съвременния строителен софтуер се състои в това, че не автоматизацијата или повишената ефективност, а способността на инженерите да разбират конкретен високоспециализиран софтуер влияе върху качеството и цената на обработката на данните за строителните проекти, както и върху печалбите и дългосрочното оцеляване на компаниите, които изпълняват строителни проекти.

Липсата на достъп до бази данни CAD -системи, които се използват в десетки други системи и стотици процеси [63], и произтичащата от това липса на качествена комуникация между отделните специалисти доведоха строителната индустрия до статута на един от най-неefективните сектори на икономиката по отношение на производителността [44].

През последните 20 години на CAD- (BIM-) приложения за проектиране, появата на нови системи (ERP), нови строителни технологии и материали, производителността на цялата строителна индустрия е спаднала с 20% (фигура 2.2-1), докато общата производителност на всички сектори на икономиката, които нямат сериозни проблеми с достъпа до бази данни и маркетингови концепции, подобни на BIM, се е увеличила със 70% (96% в производствената индустрия) [122].



Фиг. 6.2-1 Поради изолираността и сложността на проектните данни, от които зависят десетки отдели и стотици процеси в строителството, скоростта на вземане на решения е няколко пъти по-бавна, отколкото в други отрасли.

Съществуват обаче и отделни примери за алтернативни подходи за създаване на оперативна съвместимост между CAD решения. Най-голямата строителна компания в Европа с проекта SCOPE [123], започнал още през 2018 г., демонстрира как е възможно да се излезе извън класическата логика на CAD- (BIM-) системите. Вместо да се опитват да подчинят IFC или да разчитат на патентованите геометрични ядра, разработчиците на SCOPE използват API и SDK за обратно инженерство, за да извличат данни от различни CAD програми, да ги преобразуват в неутрални формати като OBJ или CPIXML на базата на единственото геометрично ядро с отворен код OCCT, и по-нататък да ги прилагат в стотици бизнес процеси на строителни и проектантски компании. Въпреки напредничавостта на идеята обаче, такива проекти се сблъскват с ограниченията и сложността на свободните геометрични ядра и все още остават част от затворени екосистеми на една компания, които възпроизвеждат логиката на едноличните решения.

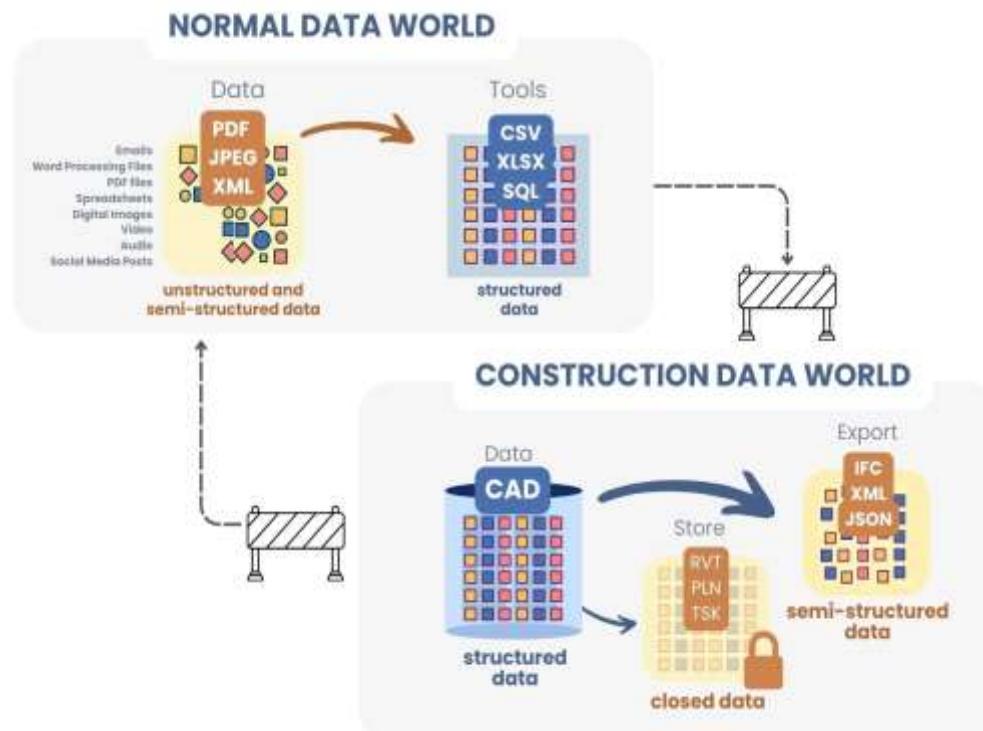
Поради ограниченията на затворените системи и различията във форматите на данни, както и поради липсата на ефективни инструменти за тяхното обединяване, компаниите, които трябва да работят с CAD формати, се сблъскват с натрупването на значителни количества данни с различна степен на структура и затвореност. Тези данни не се използват правилно и изчезват в архивите, където остават завинаги забравени и неизползвани.

Данните, получени със значителни усилия във фазата на проектиране, стават недостъпни за по-нататъшно използване поради своята сложност и затворен характер.

В резултат на това през последните 30 години разработчиците в строителния бранш са принудени да се сблъскват с един и същ проблем отново и отново: всеки нов затворен формат или патентовано решение поражда необходимостта от интеграция със съществуващите отворени и затворени CAD системи. Тези постоянни опити за осигуряване на оперативна съвместимост между различните CAD и BIM решения само усложняват екосистемата от данни, вместо да допринасят за нейното опростяване и стандартизиране.

## Митът за оперативната съвместимост между CAD системите

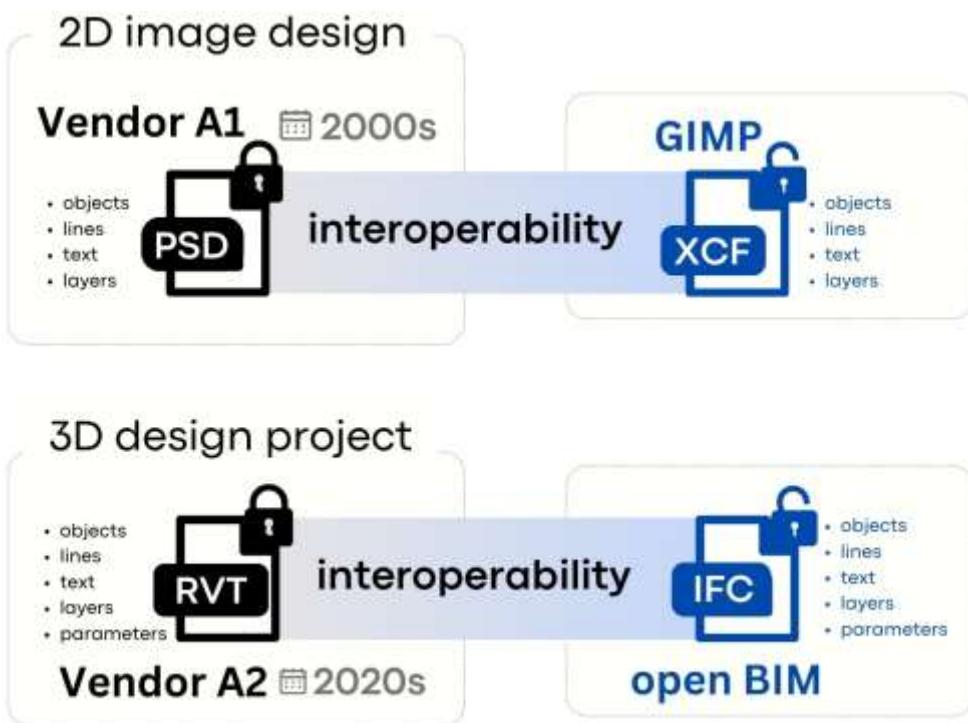
Ако в средата на 90-те години на миналия век основната насока на развитието на оперативната съвместимост в CAD средата беше разчупването на патентования формат DWG - което завърши с победата на алианса Open DWG [75] и действителното отваряне на най-популярния формат за чертежи за цялата строителна индустрия - то в средата на 2020-те години фокусът се измества. Нова тенденция набира скорост в строителната индустрия: множество екипи от разработчици се фокусират върху създаването на така наречените "мостове" между затворените CAD системи (closed BIM), IFC формата и отворените решения (open BIM). Повечето от тези инициативи се основават на използването на формата IFC и геометричното ядро OCCT, като осигуряват технически мост между различни платформи. Този подход се разглежда като обещаващо направление, което може значително да подобри обмена на данни и оперативната съвместимост на софтуерните инструменти.



**Фигура 6.2-2** Докато други индустрии работят с отворени данни, строителната индустрия трябва да работи със затворени или слабо структурирани CAD формати (BIM).

Подобен подход има исторически паралели. През 2000 г. разработчиците, опитвайки се да преодолеят господството на най-големия доставчик на графични редактори (2D world), се опитват да създадат безпроблемна интеграция между своето патентовано решение и безплатния софтуер с отворен код - алтернатива на GIMP (фиг. 6.2-3). Тогава, както и днес в строителството, ставаше въпрос за опит да се свърже затворената и отворената система, като се запазят сложните параметри, слоеве и вътрешната логика на софтуера.

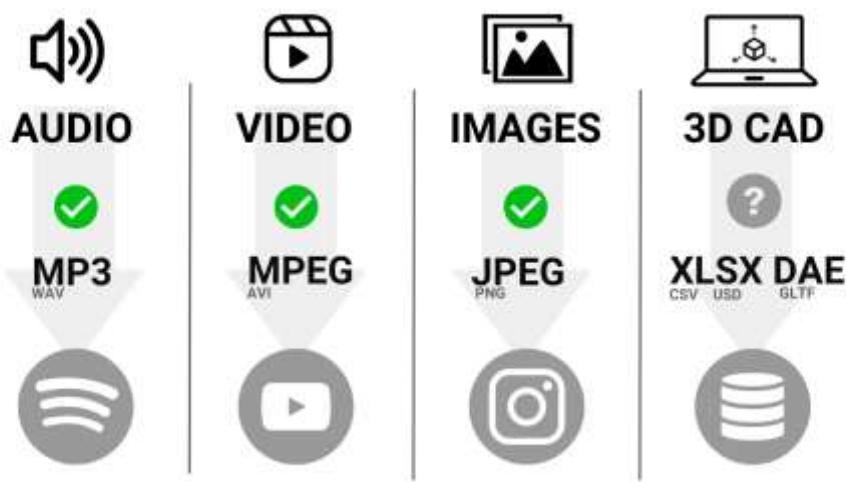
Потребителите обаче всъщност търсеха прости решения - плоски, отворени данни без прекомерна сложност на слоеве и програмни параметри (аналог на геометричното ядро в CAD). Потребителите търсеха прости и отворени формати за данни, без излишна логика. В графиката такива формати станаха JPEG, PNG и GIF. Днес те се използват в социалните мрежи, на уебсайтове, в приложения - лесно се обработват и интерпретират, независимо от платформата или производителя на софтуера.



**Фигура 6.2-3** Оперативната съвместимост на форматите за данни в строителството е подобна на пътя от опитите за обединяване на патентован продукт на популярен производител и GIMP с отворен код през 2000-те години.

В резултат на това днес почти никой в индустрията за обработка на изображения не използва затворени формати като PSD или отворен XCF за приложения, социални мрежи като Facebook и Instagram или като съдържание в уебсайтове. Вместо това при повечето задачи се използват плоски и отворени формати JPEG, PNG и GIF заради лесната употреба и широката съвместимост.

Отворените формати като JPEG и PNG се превърнаха в стандарт за споделяне на изображения поради своята гъвкавост и широка поддръжка, което ги прави лесни за използване на различни платформи. Подобен преход може да се наблюдава и при други формати за обмен, като видео и аудио, където се наблюга на универсални формати като MPEG и MP3 заради тяхната ефективност на компресия и широка съвместимост. Подобен преход към стандартизация опростява обмена и възпроизвеждането на съдържание и информация, като ги направи достъпни за всички потребители на различни платформи (фигура 6.2-4).



Фигура 6.2-4 Опростените формати без сложни функции за редактиране станаха популярни за споделяне и използване на данни.

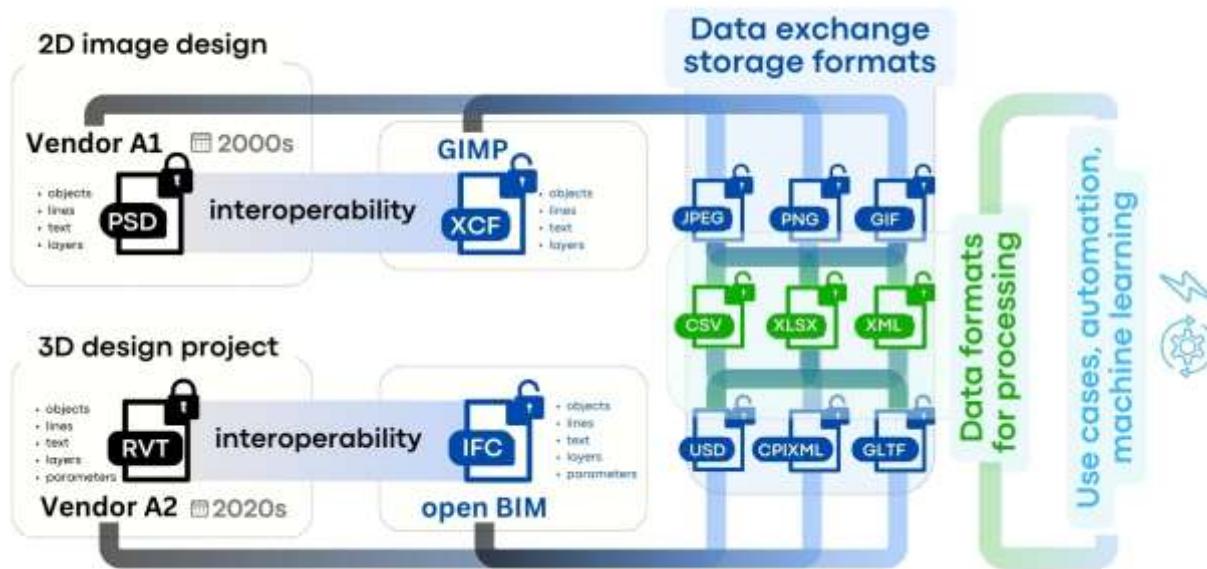
Подобни процеси протичат и при 3D моделирането на. Прости и отворени формати като USD, OBJ, glTF, DAE, DXF, SQL и XLSX все по-често се използват в проектите за обмен на данни извън CAD средата (BIM). В тези формати се съхранява цялата необходима информация, включително геометрия и метаданни, без да е необходимо да се работи със сложна BREP структура, геометрични ядра или специфични за доставчика вътрешни класификатори. Собствени формати като NWC, SVF, SVF2, CPIXML и CP2, предоставяни от водещи производители на софтуер, също изпълняват подобни функции, но остават затворени, за разлика от отворените стандарти.

Заслужава да се отбележи (и си струва да се припомни отново, както вече беше споменато в предишната глава), че тази идея - отхвърлянето на междинни неутрални и параметрични формати като IGES, STEP и IFC - беше подкрепена още през 2000 г. от най-големия доставчик на CAD, който създаде BIM Whitepaper и регистрира формата IFC през 1994 г. В бялата книга от 2000 г. "Интегрирано проектиране и производство" [65] производителят на CAD подчертава значението на естествения достъп до CAD базата данни в рамките на софтуерната среда, без да е необходимо да се използват междинни транслятори и параметрични формати, за да се запази пълнотата и точността на информацията.

Строителният бранш все още не е постигнал съгласие нито по отношение на инструментите за достъп до базите данни на CAD или тяхното принудително обратно проектиране, нито по отношение на приемането на общ опростен формат на данните за използване извън платформите на CAD

(BIM). Например, много големи компании в Централна Европа и немскоговорящите региони, работещи в строителния сектор, използват формата CPIXML в своите ERP -системи [121]. Този патентован формат, който е вид XML, обединява данни за CAD (BIM) проекти, включително геометрични и метаданни, в една организирана опростена структура. Големите строителни компании също създават свои собствени нови формати и системи, както в проекта SCOPE, който разглеждахме в предишната глава.

Затворената логика на параметричните CAD формати или сложните параметрични файлове IFC (STEP) са излишни в повечето бизнес процеси. Потребителите търсят опростени и плоски формати, като USD, CPIXML, XML &OBJ, DXF, glTF, SQLite, DAE &XLSX, които съдържат цялата необходима информация за елементите, но не са обременени с излишна геометрична логика на BREP, зависимост от геометрични ядра и вътрешни класификации на специфични CAD и BIM -продукти (фиг. 6.2-5).



Фиг. 6.2-5 За повечето случаи на употреба потребителите избират възможно най-простите формати, които са независими от програмите на доставчика.

Появата на плоски формати за изображения като JPEG, PNG и GIF, освободени от излишната логика на вътрешните двигатели на производителите, стимулира разработването на хиляди оперативно съвместими решения за обработка и използване на графики. Това доведе до появата на приложения, вариращи от инструменти за ретуширане и филтриране до социални мрежи като Instagram, Snapchat и Canva, където тези опростени данни могат да се използват, без да се обвързват с конкретен разработчик на софтуер.

Стандартизацията и опростяването на CAD форматите за проектиране ще стимулира появата на много нови, удобни за ползване и независими инструменти за работа със строителни проекти.

Преминаването от сложната логика на приложенията на доставчиците, обвързани със затворени геометрични ядра, към универсални отворени формати, базирани на библиотеки от опростени

елементи, създава предпоставки за по-гъвкава, прозрачна и ефективна обработка на данните. Това също така отваря достъп до информация за всички страни, участващи в строителния процес - от проектантите до клиентите и службите за поддръжка.

Въпреки това е много вероятно през следващите години доставчиците на CAD да се опитат отново да изместят дебата за оперативната съвместимост и достъпа до CAD бази данни. Вече ще става въпрос за "нови" концепции - като гранулирани данни, интелигентни графики, "обединени модели", цифрови близнаци в облачни хранилища - както и за създаване на индустриални съюзи и стандарти, които да продължат пътя на BIM и отворения BIM. Въпреки атрактивната терминология, подобни инициативи могат отново да се превърнат в инструменти за задържане на потребителите в рамките на собственически екосистеми. Пример за това е активното популяризиране на формата USD (Universal Scene Description) като "нов стандарт" за междуплатформено сътрудничество в областта на CAD (BIM) от 2023 г.

## Към USD и гранулирани данни

Появата на алианса AOUSD [124] през 2023 г. бележи важен обрат в строителната индустрия. Свидетели сме на началото на една нова реалност, оформена от доставчиците на CAD, в работата със строителни данни чрез няколко значителни промени. Първата голяма промяна е свързана с възприемането на CAD-данните. Специалистите, участващи в ранните етапи на концептуалното проектиране, все повече осъзнават, че създаването на проект в CAD среда е само начална точка. Данните, генериирани в процеса на проектиране, в крайна сметка се превръщат в основа за анализ, експлоатация и управление на обектите. Това означава, че те трябва да бъдат достъпни и използваеми в системи извън традиционните CAD инструменти.

Успоредно с това се извършва революция в подхода на водещите разработчици. Водещият доставчик на CAD-, създател на концепцията за BIM и на формата IFC, прави неочекван завой в стратегията си. От 2023 г. компанията се отказва от традиционното съхранение на данни в отделни файлове, като се фокусира върху работата с гранулирани (normalизирани и структурирани) данни и преминава към подход, ориентиран към данните [125].

Доставчиците следват историческите тенденции в други индустрии: повечето потребители не се нуждаят от затворени CAD формати (подобни на PSD) или сложни параметрични IFC файлове (подобни на GIMP с логика на слоевете). Те се нуждаят от прости изображения на обекти, които могат да се използват в CAFM (строителен Instagram), ERP (Facebook) и хиляди други процеси, изпълнени с електронни таблици в Excel и документи в PDF.

Съвременните тенденции в строителната индустрия потенциално създават предпоставки за постепенно изместване на параметричните и сложни формати в полза на по-универсалните и независими формати USD, GLTF, DAE, OBJ (с метаинформация както в рамките на хибрида, така и в отделни структурирани или свободно структурирани формати). Историческите лидери, включително големите проектантски компании, които някога активно популяризираха IFC в средата на 90-те години, сега открито популяризират новия формат USD [93], като подчертават неговата простота и универсалност (фиг. 6.2-6). Масовото внедряване на USD в продуктите, съвместимостта с GLTF и активната интеграция в инструменти като Blender, Unreal Engine и Omniverse показват потенциала за

начало на нова парадигма за работа с данни. Наред с популярността на локализирани решения като европейския плосък формат на USD - CPIXML, използван в популярни европейски ERP, може потенциално да засили позициите на USD в Централна Европа. Организациите, участващи в разработването на формата IFC, вече адаптират своята стратегия към USD [126], което само потвърждава неизбежността на промяната.

Technical Specifications				Comparison / Notes
File Structure	Monolithic file	Uses ECS and linked data	IFC stores all data in one file; USD uses Entity-Component-System and linked data for modularity and flexibility	
Data Structure	Complex semantics, parametric geometry	Flat format, geometry in MESH, data in JSON	IFC is complex and parametric; USD is simpler and uses flat data	
Geometry	Parametric, dependent on BREP	Flat, MESH (triangular meshes)	IFC uses parametrics; USD uses meshes for simplified processing	
Properties	Complex structure of semantic descriptions	Properties in JSON, easy access	Properties in USD are easier to use thanks to JSON	
Export/Import	Complex implementation, dependent on third-party SDKs	Easy integration, wide support	USD integrates more easily and is supported in many products	
Format Complexity	High, requires deep understanding	Low, optimized for convenience	The time required to understand the structure of the file and the information stored in it.	
Performance	Can be slow when processing large models	High performance in visualization and processing	USD is optimized for speed and efficiency. Simulations, machine learning, AI, smart cities will be held in the Nvidia Omniverse	
Integration with 3D Engines	Limited	High, designed for graphics engines	USD excels with native support for real-time visualization platforms	
Support outside CAD Software	BlenderBIM, IfcOpenShell	Unreal Engine, Unity, Blender, Omniverse	USD is widely supported in graphics tools	
Cloud Technology Support	Limited	Well-suited for cloud services and online collaboration	USD is optimized for cloud solutions	
Ease of Integration into Web Applications	Difficult to integrate due to size and complexity	Easy to integrate, supports modern web technologies	USD is preferable for web applications	
Change Management	Versions through separate files	Versioning built into the format core	IFC handles changes via separate files, while USD embeds versioning directly into its structure	
Collaboration Support	Supports data exchange between project participants	Designed for collaborative work on complex scenes	USD provides efficient collaboration through layers and variations	
Learnability	Steep learning curve due to complexity	Easier to master thanks to a clear structure	USD is easier to learn and implement	

Фигура 6.2-6 Сравнение на техническите спецификации на форматите IFC и USD.

На този фон USD има потенциала да се превърне във фактически стандарт, като обещава да преодолее много от настоящите ограничения, свързани най-вече със сложността на съществуващите CAD - (BIM-) формати и зависимостта на тяхното тълкуване от геометрични ядра.

Вместо параметричните и сложни CAD -формати и IFC - опростените формати за данни USD, gLTF, DAE, OBJ с метаинформация за елементите в CSV, XLSX, JSON, XML ще заемат място в строителната индустрия поради своята простота и гъвкавост.

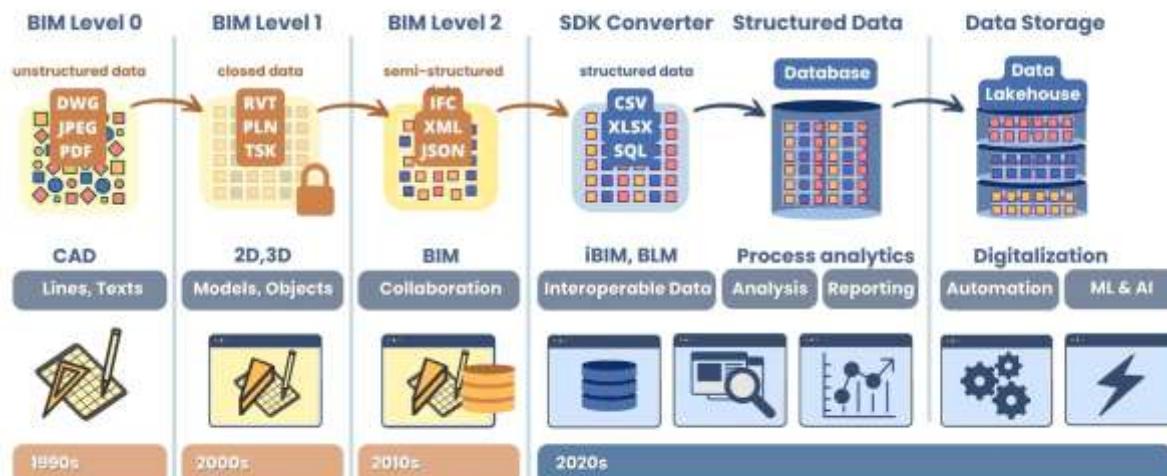
Настоящите промени в строителния бранш на пръв поглед изглеждат като технологичен пробив, свързан с прехода от останалия IFC към по-modерния USD. Струва си обаче да се има предвид, че още през 2000 г. същият производител на CAD, който е разработил IFC, пише за неговите проблеми и необходимостта от достъп до базата данни [65], а сега активно насърчава прехода към нов стандарт - USD.

Зад поредната фасада на "отворените данни" USD и "новите" концепции за управление на гранулирани данни чрез базирани в облака приложения, които доставчиците на CAD започват да популяризират, може да се крие намерението на доставчиците да монополизират управлението на проектните данни, при което потребителите се оказват в положение, при което изборът на формат е свързан повече с корпоративни интереси, отколкото с реални нужди.

Анализът на основните факти [93] показва, че основната цел на тези промени е по-скоро не толкова удобството на потребителите, колкото запазването на контрола върху екосистемите и потоците от данни в полза на доставчиците, които за 40 години никога не са били в състояние да предоставят достъп до базите данни на CAD.

Може би сега е моментът компаниите да престанат да чакат нови концепции от доставчиците на софтуер и да се съсредоточат върху самостоятелно развитие в посока, ориентирана към данните. След като се освободи от проблемите с достъпа до данни чрез инструментите за обратен инженеринг, индустрията ще може да премине самостоятелно към модерни, безплатни и удобни инструменти за работа и анализ на данни, без да налага нови концепции.

### CAD (BIM) Maturity Levels: From Stage 0 to Structured Data



Фиг. 6.2-7 Ниво на зрълост на CAD (BIM): от неструктурирани данни към структурирани данни и хранилища.

Достъпът до бази данни, отворените данни и формати неизбежно ще се превърнат в стандарт в строителната индустрия, независимо от опитите на доставчиците да забавят процеса - това е само въпрос на време (фиг. 6.2-7). Темпото на този преход може да се увеличи значително, ако все повече специалисти се запознаят с отворените формати, инструментите за бази данни и наличните SDK за обратен инженеринг, позволяващи директен достъп до CAD данни -системи [92].

Бъдещето е в отворените, унифицирани и аналитично достъпни данни. За да избегнат зависимостта от решенията на доставчиците и за да не станат заложници на затворени екосистеми, строителните и инженерните компании рано или късно ще трябва да заложат на отвореността и независимостта, като изберат формати и решения, които осигуряват пълен контрол върху данните.

Данните, които се създават днес в строителния бранш, ще бъдат ключов ресурс за бизнес решения в бъдеще. Те ще играят ролята на стратегическо "гориво", което ще подхранва развитието и ефективността на строителните компании. Бъдещето на строителния бранш е в способността за работа с данни, а не в избора на формати или модели на данни.

За да се разбере разликата между отворените формати USD, glTF, DAE, OBJ и патентованите параметрични CAD формати, е важно да се разгледа един от най-сложните и ключови елементи на данните при визуализация и проектни изчисления - геометрията и процесите на нейното генериране. А за да се разбере как геометричните данни стават основа за анализи и изчисления в строителството, е необходимо да се вникне по-дълбоко в механизмите на генериране, трансформиране и съхранение на геометрията.



## ГЛАВА 6.3.

### ГЕОМЕТРИЯ В СТРОИТЕЛСТВОТО: ОТ ЛИНИИ ДО КУБИЧНИ МЕТРИ

#### **Когато линиите се превръщат в пари или защо строителите се нуждаят от геометрия**

Геометрията в строителството е не само визуализация, но и основа за точни количествени изчисления. В модела на проекта геометрията допълва списъците с параметри на елементите (фиг. 3.1-16) с важни обемни характеристики като дължина, площ и обем. Стойностите на тези обемни параметри се изчисляват автоматично с помощта на геометричните ядра и са отправна точка за оценки, графици и ресурсни модели. Както вече обсъдихме в част 5 на тази книга и в главата "Остойностяване и оценяване на строителни проекти", именно обемните параметри на групите обекти от CAD -моделите са в основата на съвременните ERP, PMIS -системи/ Геометрията играе основна роля не само във фазата на проектиране, но и в управлението на изпълнението на проекта, контрола на графиците, бюджетирането и експлоатацията. Точно както преди хиляди години, при изграждането на египетските пирамиди, точността на проекта е зависела от мерките за дължина като лакът и лакът, така и днес точността на интерпретацията на геометрията в CAD -програмите влияе пряко върху резултата: от бюджета и сроковете до избора на изпълнител и логистиката на доставките

В условията на силна конкуренция и ограничен бюджет точността на обемните изчисления, която пряко зависи от геометрията, се превръща във фактор за оцеляване.

Съвременните ERP -системи зависят пряко от правилните обемни характеристики, получени от CAD - и BIM -модели. Ето защо точното геометрично описание на елементите не е просто визуализация, а ключов инструмент за управление на разходите и времето за строителство.

В миналото геометрията е била основният език на инженерната комуникация. От линиите върху папирус до цифровите модели, чертежите и геометричните изображения са служили като средство за обмен на информация между проектанти, майстори и оценители. Преди появата на компютрите изчисленията са се извършвали ръчно с помощта на линийки и транспортири. Днес тази задача е автоматизирана благодарение на обемното моделиране: геометричните ядра на CAD софтуера превръщат линиите и точките в триизмерни тела, от които автоматично се извличат всички необходими характеристики.

При работа в CAD -програми създаването на геометрични елементи за изчисления се извършва чрез потребителския интерфейс на CAD- (BIM-) програми. За трансформиране на точки и линии в обемни тела се използва геометричното ядро, което изпълнява ключовата задача - трансформиране на геометрията в обемни модели, от които след апроксимация автоматично се изчисляват обемните характеристики на елемента.

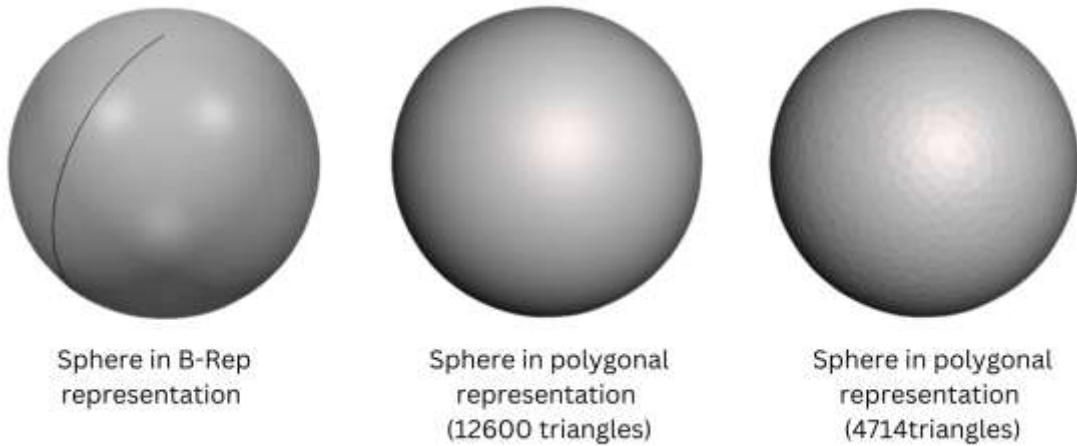
#### **От линии към обеми: как площта и обемът се превръщат в данни**

В инженерната практика обемите и площините се изчисляват от геометрични повърхности, описани

аналитично или чрез параметрични модели, като NURBS (неравномерни рационални B-сплайни) в рамките на BREP (представяне на гранични елементи).

NURBS (Non-Uniform Rational B-Splines - неравномерни рационални B-образни линии) е математически начин за описване на криви и повърхности, докато BREP е рамка за описание на пълната триизмерна геометрия на даден обект, включително неговите граници, които могат да бъдат определени с помощта на NURBS.

Въпреки точността на BREP и NURBS, те изискват мощни изчислителни ресурси и сложни алгоритми. Въпреки това директното изчисление от такива математически точни описания често е изчислително трудно, така че на практика почти винаги се използва теселиране - преобразуване на повърхностите в мрежа от триъгълници - което опростява последващите изчисления. Теселиране е разделянето на сложна повърхнина на триъгълници или многоъгълници. В средите на CAD /CAE този метод се използва за визуализация, изчисления на обеми, търсене на колизии, експортиране във формати като MESH и анализ на колизии. Пример от природата е пчелната пита, при която сложната форма е разбита на правилна мрежа (фиг. 6.3-1).



Фиг. 6.3-1 Една и съща сфера в параметрично описание BREP и полигонално представяне различен брой триъгълници.

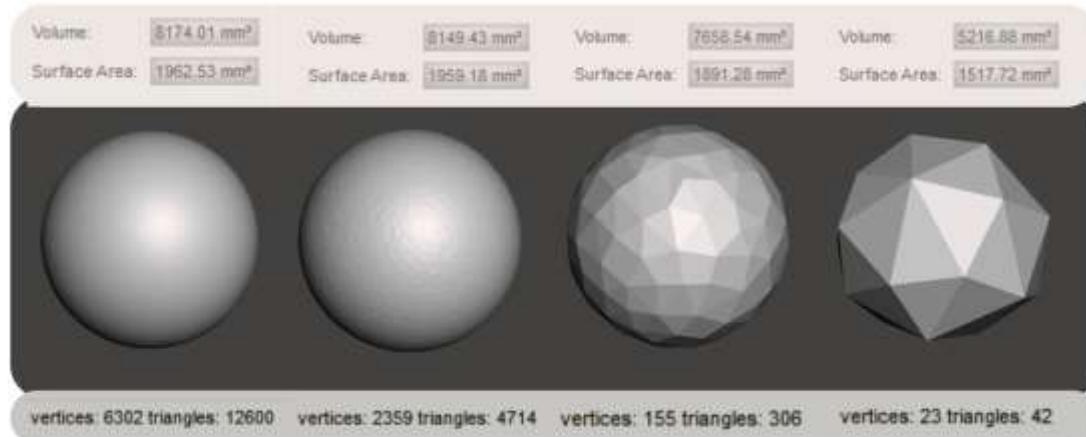
BREP (NURBS), използван в CAD, не е основен модел на геометрията. Той е създаден като удобен инструмент за представяне на кръгове и рационални сплайни и за минимизиране на съхранението на геометрични данни. Въпреки това той има ограничения - например невъзможността да се опише точно синусоидата, която е в основата на спираловидните линии и повърхности, и необходимостта да се използват сложни геометрични ядра.

За разлика от тях триъгълните мрежи и теселирането на параметрични форми се характеризират с простота, ефективно използване на паметта и възможност за обработка на големи количества данни (фиг. 6.3-2). Тези предимства позволяват при изчисляването на геометрични фигури да се мине без сложни и скъпи геометрични ядра, както и без десетките милиони редове код, вградени в тях.

В повечето случаи на сгради няма значение как точно са определени обемните характеристики - чрез параметрични модели (BREP, IFC) или чрез полигони (USD, glTF, DAE, OBJ). Геометрията си остава форма на приближение: независимо дали чрез NURBS или MESH, тя винаги е приблизително описание на формата.

Геометрията, дефинирана като полигони или BREP (NURBS), остава до известна степен само начин за приблизително описание на непрекъсната форма. Точно както интегралите на Френел нямат точен аналитичен израз, дискретизирането на геометрията чрез полигони или NURBS винаги е приближение, точно както триъгълната MESH.

Параметричната геометрия във формат BREP е необходима най-вече в случаите, когато е важен минималният размер на данните и е възможно да се използват ресурсоемки и скъпи геометрични ядра за тяхната обработка и показване. Най-често това е характерно за разработчиците на CAD-програми, които за тази цел прилагат в своите продукти геометрични ядра на MCAD-доставчици. В този случай, дори в рамките на тези програми, BREP-моделите в процеса на теселиране за визуализация и изчисления често се преобразуват в триъгълници (подобно на начина, по който PSD-файловете се опростяват в JPEG).



**Фиг. 6.3-2 Разлика в обемните характеристики при фигури с различен брой многоъгълници.**

Полигоналният MESH, както и параметричният BREP, имат своите предимства и ограничения, но целта е една и съща - да се опише геометрията с оглед на задачата на потребителя. В крайна сметка точността на един геометричен модел зависи не само от метода на неговото представяне, но и от изискванията на конкретната задача.

При повечето проблеми на конструкцията нуждата от параметрична геометрия и сложни геометрични ядра може да се окаже излишна.

При всяка конкретна задача за автоматизиране на изчисленията си струва да се помисли дали значението на параметричната геометрия не се преувеличава от разработчиците на CAD, които са заинтересовани да популяризират и продават собствените си софтуерни продукти.

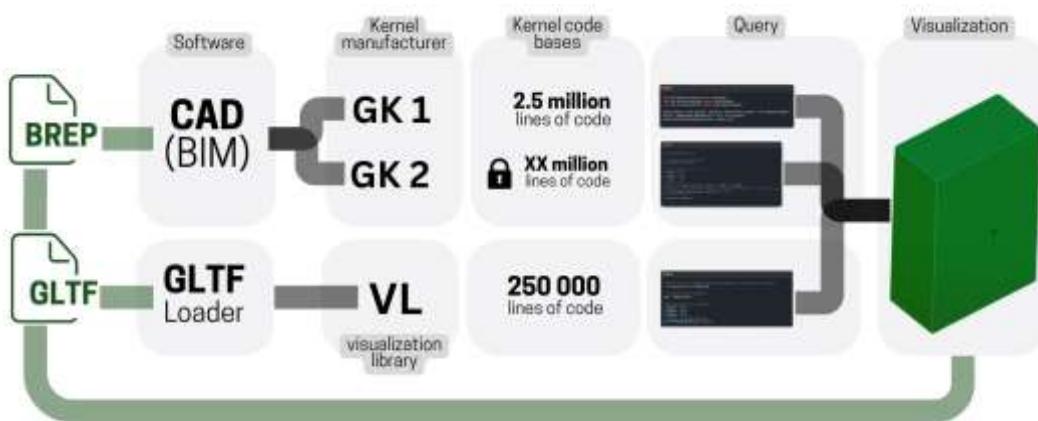
## Преминаване към MESH, USD и полигони: използване на теселиране за геометрия

В строителната индустрия, когато се стриймват, разработват системи, бази данни или се автоматизират процеси, които работят с проектна информация и геометрия на елементи, е важно да се търси независимост от конкретни CAD редактори и геометрични ядра.

Форматът за обмен, който ще се използва както в отделите за изчисления, така и на строителната площадка, не трябва да се основава на конкретна CAD- (BIM-) програма. Геометричната информация следва да бъде представена във формата директно чрез теселиране, без позоваване на геометричното ядро или CAD архитектурата.

Параметричната геометрия от CAD може да се разглежда като междинен източник, но не и като основа на универсален формат. Повечето параметрични описания (включително BREP и NURBS) при всички случаи се преобразуват в полигонален MESH за по-нататъшна обработка. Ако резултатът е един и същ (теселиране и полигони) и процесът е по-прост, изборът е очевиден. Това е подобно на избора между графови онтологии и структурирани таблици (който разгледахме в четвърта част): прекомерната сложност рядко е оправдана (фиг. 3.2-10, фиг. 6.1-8).

Отворените формати, като например: OBJ, STL, glTF, SVF, CPIXML, USD и DAE, използват универсална триъгълна мрежова структура, което им дава значителни предимства. Тези формати имат отлична оперативна съвместимост - те са лесни за четене и визуализиране с помощта на налични библиотеки с отворен код, без да е необходимо да се използват сложни специализирани геометрични ядра, съдържащи милиони редове код (фигура 6.3-3). Тези универсални геометрични формати се използват в приложения, вариращи от сравнително прости инструменти за проектиране на кухни в IKEA™ до сложни системи за визуализация на обекти в киното и VR -приложения. Важно предимство е наличието на голям брой бесплатни библиотеки с отворен код за работа с тези формати, достъпни за повечето платформи и езици за програмиране.



Фигура 6.3-3 Същото представяне на геометрията се постига чрез използване на параметрични формати и геометрични ядра, или чрез използване на триъгълни формати и библиотеки за визуализация с отворен код.

Освен самите потребители, доставчиците на CAD се сблъскват с проблеми при интерпретирането на чужди параметрични CAD формати или отворени IFC поради различните геометрични ядра. На практика всички CAD -доставчици, без изключение, използват SDK за обратно инженерство за прехвърляне на данни между системите и никой от тях не разчита на формати като IFC или USD [93] за целите на оперативната съвместимост.

Вместо да се използват концепции, настъчавани от съюзи на доставчици на CAD-, които те самите не използват разработчиците и потребителите на CAD решения е по-продуктивно да се съсредоточат върху разбирането на ползите от всеки подход в конкретен контекст и да изберат един или друг вид геометрия в зависимост от случая на използване. Изборът между различни геометрични представления е компромис между точността, изчислителната ефективност и практическите нужди на конкретната задача.

Сложността, свързана с използването на геометрични ядра, традиционно налагана на строителния бранш от големите доставчици при обработката на проектни данни, често се оказва излишна. Форматът USD, базиран на геометрията MESH, може да се превърне в своеобразна "кутия на Пандора" за индустрията, откривайки нови възможности за разработчиците да организират обмена на данни - извън рамките на IFC и параметричните структури BREP, типични за CAD доставчиците.

След по- внимателно разглеждане на структурата на USD, DAE, glTF, OBJ и т.н. става ясно, че съществуват по-прости, отворени формати, които позволяват ефективно да се организира прехвърлянето и използването на геометрична информация, без да е необходимо да се разчита на сложни параметри и затворени геометрични ядра. Този подход не само понижава техническия праг за навлизане на разработчиците, но и благоприятства разработването на гъвкави, мащабиращи и наистина отворени решения за цифрово строителство.

## **LOD, LOI, LOMD - уникална класификация на детайлите в CAD (BIM)**

В допълнение към форматите за геометрично представяне, в свят, в който различните индустрии използват различни нива на детайлност и дълбочина на данните, методологиите на CAD - (BIM-) предлагат свои собствени уникални класификационни системи, които структурират подхода към информационните модели на сградите.

Един от примерите за нови подходи към стандартизацията е въвеждането на нива на разработване на модели, които отразяват степента на готовност и надеждност на графичните и информационните компоненти. За диференциране на информационното съдържание при работа с CAD - (BIM-) данни се появиха LOD (Level Of Detail) - ниво на детайлност на графичната част на модела, и LOI (Level Of Information) - ниво на разработка на данните. Освен това за интегрирания подход беше въведено понятието LOA (Level of Accuracy) - точността на представените елементи и LOG (Level of Geometry) за определяне на точността на графичното представяне.

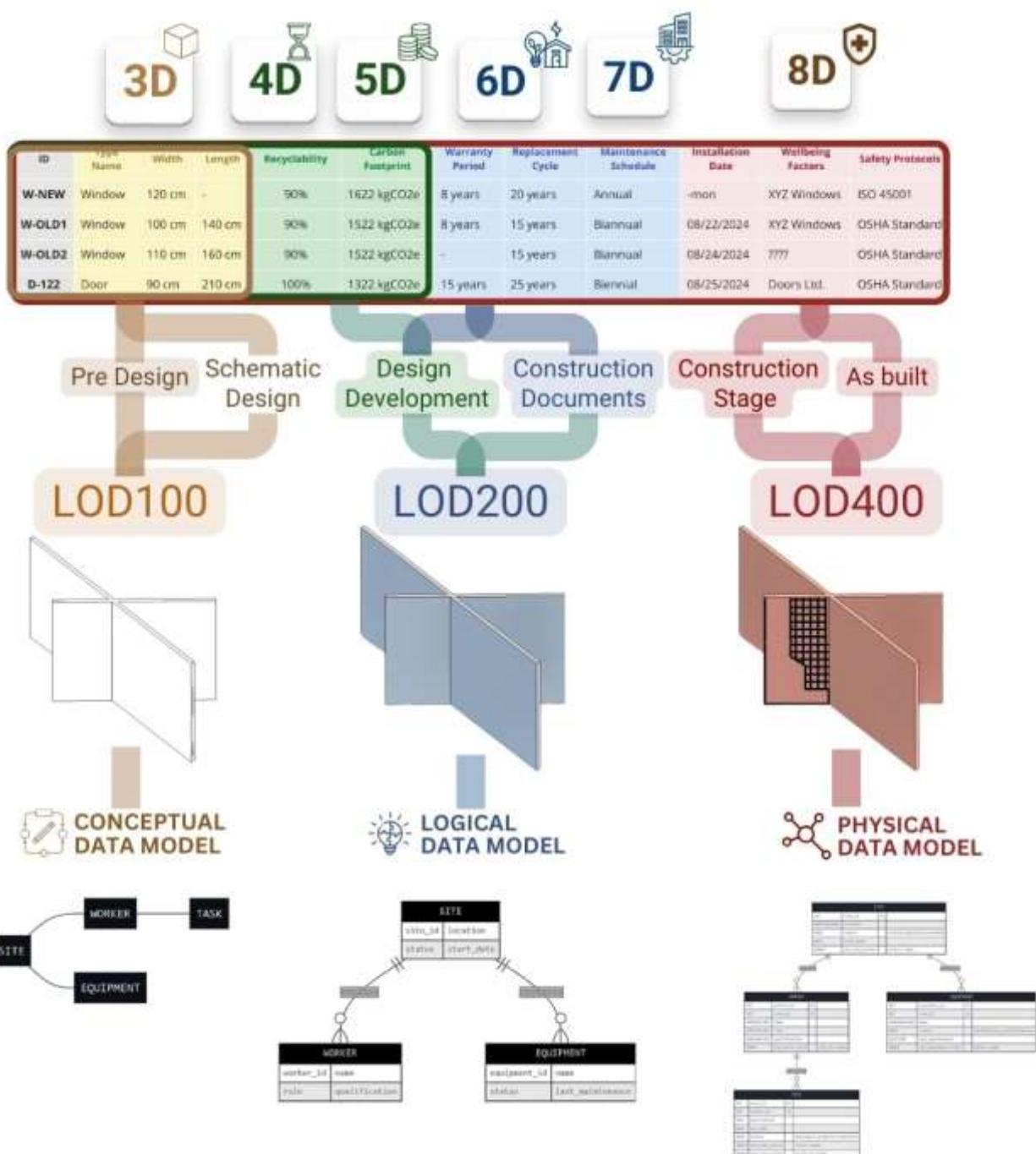
Нивата на детайлност (LOD) са обозначени с числа от 100 до 500, които отразяват степента на разработване на модела. LOD 100 е концептуален модел с общи форми и размери. LOD 200 включва по-точни размери и форми, но с условна детайлност. LOD 300 е подробен модел с точни размери, форми и разположение на елементите. LOD 400 съдържа подробна информация, необходима за

изработката и монтажа на елементите. LOD 500 отразява действителното състояние на съоръжението след изграждането му и се използва за експлоатация и поддръжка. Тези нива описват структурата на насищане с информация на CAD (BIM) моделите на различни етапи от жизнения цикъл, включително 3D, 4D, 5D и след това.

В реалните проекти високото ниво на детайлност (LOD400) често е прекомерно и е достатъчно да се използва геометрия LOD100 или дори плоски чертежи, докато останалите данни могат да бъдат получени или по изчислителен път, или от свързани елементи, които може да нямат ясно изразена геометрия. Например пространствата и елементите на помещенията (категориите елементи на помещанията) може да нямат визуална геометрия, но съдържат значителни количества информация и бази данни, около които са изградени много бизнес процеси.

Ето защо е важно ясно да се определи необходимото ниво на детайлност преди започване на проектирането. За случаите на използване на 4D -7D често са достатъчни дори DWG чертежи и минимална геометрия LOD100. Основната задача в процеса на изискванията е да се намери баланс между богатството и практичността на модела.

По същество, ако разглеждаме CAD (BIM) данните като база данни (а те са такива), описанietо на насищането с модели чрез нови съкращения не е нищо друго освен поетапно моделиране на данни за информационни системи, като се започне от концептуалното ниво и се стигне до физическото (фиг. 6.3-4), което беше разгледано подробно в трета и четвърта част на книгата. Всяко увеличаване на LOD и LOI означава добавяне на информация, необходима за нови задачи: изчисления, управление на строителството, експлоатация, и се характеризира с последователно обогатяване на модела с допълнителни информационни слоеве (3D -8D) под формата на различни параметри, които разглеждахме в петата част на книгата.



### Фигура 6.3-4 Процесът на конкретизиране на детайлите на проекта е идентичен с моделирането на данни от концептуален към физически модел на данни.

Геометрията е само част от проектните данни, чиято необходимост невинаги е оправдана в строителните проекти, а основният проблем при работата с CAD-данни е не толкова как се визуализират моделите, а по-скоро как данните от тези модели могат да се използват извън CAD- (BIM-) програмите.

Към средата на 2000-те години строителната индустрия се изправи пред безпрецедентно предизвикателство, свързано с бързото увеличаване на обема на данните в системите за управление и обработка на данни, особено тези, които идват от CAD отделите (BIM). Това драматично нарастване на обема на данните изненада ръководителите на компаниите и те не бяха подгответи за нарастващите изисквания към качеството и управлението на данните.

## Нови CAD стандарти (BIM) - AIA, BEP, IDS, LOD, COBie

Възползвайки се от липсата на свободен достъп до CAD бази данни и ограничената конкуренция на пазара за обработка на данни, както и от маркетинговите кампании, свързани с новото съкращение BIM, организациите, участващи в разработването на подходи за работа с CAD данни, започнаха да създават нови стандарти и концепции, които де юре трябва да бъдат насочени към подобряване на практиките за управление на данни.

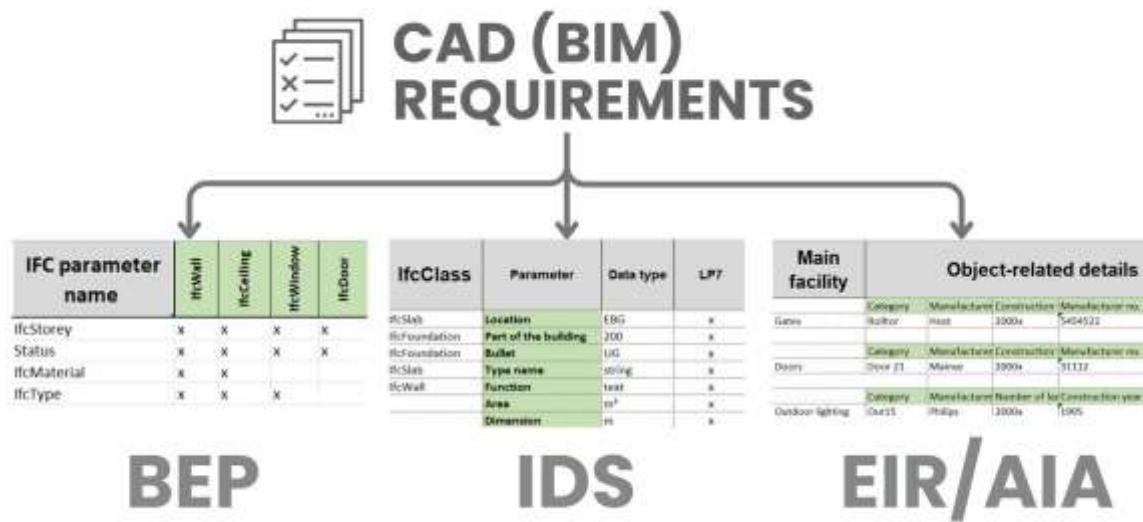
Въпреки че почти всички инициативи, пряко или непряко подкрепени от доставчиците на CAD и разработчиците (BIM), са насочени към оптимизиране на работните процеси, те доведоха до множество стандарти, лобирани от различни заинтересовани страни, което доведе до известна неяснота и объркване в строителната индустрия по отношение на процесите на предаване на данни.

Нека да изброим някои от новите стандарти за данни, освен LOD, LOI, LOA, LOG, които се появиха през последните години в строителната индустрия:

- **BEP** (BIM Execution Plan) - описва как да се интегрира и използва CAD (BIM) в даден проект, като определя методите и процесите за обработка на данни.
- **Документ EIR /AIA** (Изисквания за информация на възложителя) - изготвя се от възложителя преди обявяването на търга и съдържа изискванията към изпълнителя за подготовка и предоставяне на информация. Той служи като основа за изготвяне на BEP в съответния проект.
- **AIM** (Информационен модел на активите) е част от процеса на BIM. След като проектът бъде предаден и завършен, моделът на данните се нарича Информационен модел на активите или AIM. Целта на AIM е да управлява, поддържа и експлоатира реализирания актив.
- **IDS** (Information Delivery Specification) - определя изискванията на и какви данни и в какъв формат се изискват на различните етапи от строителния проект.
- **iLOD** е нивото на детайлност LOD, с което информацията е представена в BIM -модела. То определя колко подробна и пълна е информацията в модела - от основните геометрични представления до подробните спецификации и данни.
- **eLOD** - LOD ниво на детайлност на отделните елементи в CAD модел (BIM). То определя степента, в която е моделиран всеки елемент, и свързаната с него информация, като размери, материали, експлоатационни характеристики и други подходящи атрибути.
- **APS** (Platform Services) и други продукти на основните доставчици на CAD (BIM) - описват инструментите и инфраструктурата, необходими за създаване на свързани и отворени модели на данни.

Въпреки че деклариранията цел на въвеждането на CAD (BIM) стандарти - като LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, IDS, iLOD, eLOD - е да се подобри качеството на управление на данните и да се разширят възможностите за автоматизация, на практика използването им често води до прекомерна сложност и фрагментация на процесите. Ако разглеждаме CAD (BIM) модела като база данни, става ясно, че много от тези стандарти дублират отдавна утвърдени и ефективни подходи, използвани в други отрасли при работа с информационни системи. Вместо да опростяват и унифицират, подобни инициативи често създават допълнителна терминологична тежест и възпрепятстват прилагането на истински отворени и гъвкави решения.

Забележително е, че много от тези нови концепции всъщност заместват процесите на моделиране и валидиране на данни, които бяха подробно разгледани в първите части на книгата и които отдавна се използват в други сектори на икономиката. От друга страна, в строителството процесът на стандартизация често се движи в обратна посока - създават се нови формати за описание на данни, нови стандарти и нови концепции за валидиране на данни, които не винаги водят до реално уеднаквяване и практическа приложимост. В резултат на това, вместо да се опрости и автоматизира обработката, отрасът се сблъска с допълнителни нива на регуляция и бюрокрация (фиг. 6.3-1), което не винаги допринася за повишаване на ефективността.



Фиг. 6.3-1 Изискванията към съдържанието на данните и информацията се свеждат до описание на атрибутите и техните гранични стойности, описани с помощта на таблици.

Вместо да опростяват обработката на данни, новите концепции, свързани с CAD (BIM) данните, най-често пораждат допълнителни усложнения и спорове още на етапа на тълкуване и основни дефиниции.

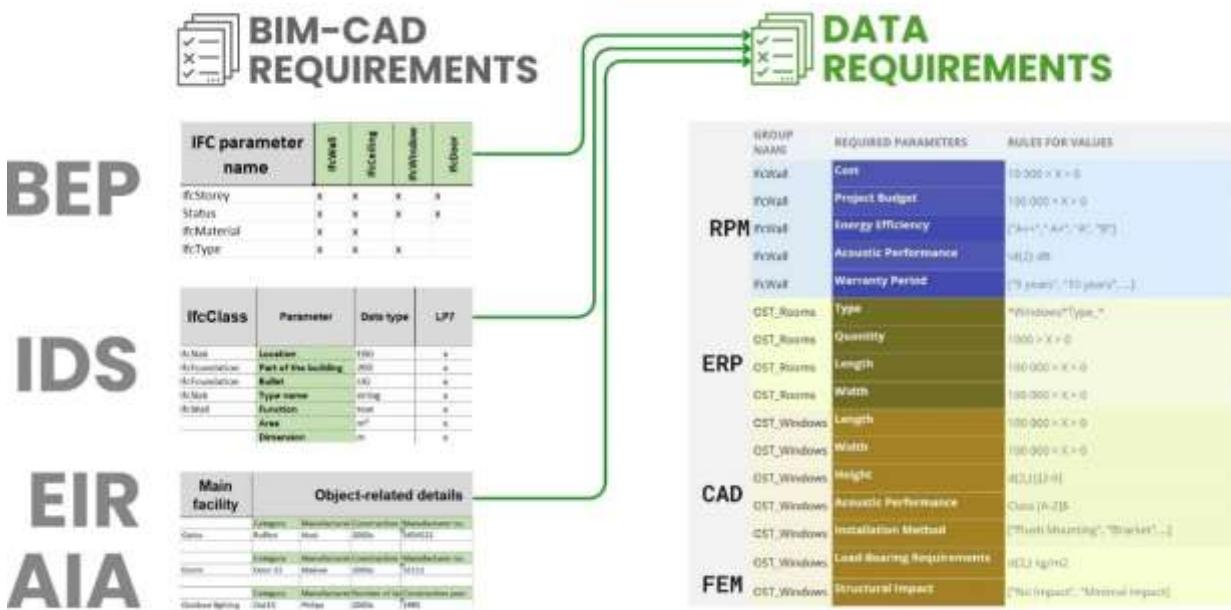
Един от последните примери за нови концепции е форматът IDS (въведен през 2020 г.), който позволява да се опишат изискванията за състава на атрибутите на информационния модел в отворената концепция BIM. Изискванията на IDS описват информация за атрибутите и техните гранични стойности под формата на структурирана таблица (Excel или MySQL), която след това се превежда

в маркировката на полуструктурни XML формат, преименуван от XML на специалното съкращение IDS.

Противно на мнението, лансирано от доставчиците и подкрепяно от BIM и отворения BIM, че обработката на данни в строителството е уникална поради използването на специализирани инструменти като CAD и BIM, форматите на данни и практиките за управление на данни в този отрасъл не се различават от тези други отрасли.

Броят на изискванията за CAD проектите и форматите (BIM) може да бъде опростен чрез използване на една таблица с изисквания и колони с атрибути, описани подробно в главата "Превеждане на изискванията в структуриран вид", без да се налага да се превеждат първоначално структурираните изисквания във формати, които не са таблични (IDS първоначално е описана чрез таблица).

Опростеният подход (Фигура 6.3-2), който включва колони за идентификатори на обекти, свойства и гранични стойности, които бяха разгледани подробно в предишни глави (Фигура 4.4-9, Фигура 4.4-16, Фигура 7.3-10), премахва необходимостта от преобразуване на изискванията във формат IDS-XML. Този метод осигурява директен, по-малко тромав и по-прозрачен механизъм за контрол на качеството на данните. Той разчита на широко използвани инструменти - от регулярни изрази (RegEx) до рамки за данни, Pandas и стандартни ETL-плейнс - точно като тези, които се използват от специалистите в други сектори на икономиката за работа с данни.



Фиг. 6.3-2 Изискванията към данните в други отрасли са опростени до структурирано описание на атрибутите и техните гранични стойности.

С течение на времето в строителния бранш, поради затворения характер на данните, се появяват все повече нови подходи и техники за контрол и управление на тези разнообразни данни, въпреки че данните в строителните проекти по същество са същите като в други области. Докато други индустрии успешно се справят със стандартизиранi подходи за обработка на данни, строителството

продължава да разработва нови и уникални формати за данни, изисквания и концепции за валидиране.

Методите и инструментите, използвани за събиране, подготовка и анализ на данни в строителството, не трябва да се различават съществено от тези, използвани от специалистите в други сектори на икономиката.

Индустрията е разработила отделна терминологична екосистема, която изисква критично осмисляне и преоценка:

- Форматът STEP е позициониран под новото име IFC, допълнено с категоризация на конструкциите, без да се вземат предвид ограниченията на самия формат STEP.
- Параметричният формат IFC се използва в процесите на предаване на данни въпреки липсата на единно геометрично ядро, необходимо за визуализация и изчисления.
- Достъпът до бази данни CAD -системи се популяризира под термина "BIM", без да се обсъждат спецификите на тези бази данни и достъпът до тях.
- Доставчиците насярчават оперативната съвместимост чрез форматите IFC и USD, често без да ги прилагат на практика, използвайки скъпоструващ обратен инженеринг, с който самите те се сблъскват трудно.
- Термините LOD, LOI, LOA, LOG, BEP, EIR, AIA, AIM, iLOD, eLOD се използват универсално за описание на едни и същи параметри на обекта, без да се правят препратки към инструментите за моделиране и проверка, които отдавна се използват в други отрасли.

Строителният бранш показва, че всичко гореизброено, макар и да звуци странно, е възможно в строителния бранш - особено ако основната цел е да се печели от всеки етап на обработката на данни чрез продажбата на специализирани услуги и софтуер. От бизнес гледна точка в това няма нищо лошо. Дали обаче подобни съкращения и подходи, свързани с CAD (BIM), наистина добавят стойност и опростяват професионалните процеси, остава отворен въпрос.

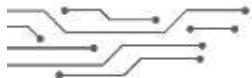
В строителния бранш такава система работи, защото самият бранш реализира по-голямата част от спекулативните си печалби в този лабиринт от системи и съкращения. Компаниите, които се интересуват от прозрачни процеси и отворени данни, са рядкост. Тази сложна ситуация вероятно ще продължи до безкрай - докато клиентите, възложителите, инвеститорите, банките и частните капитали не започнат да изискват по-ясни и по-информирани подходи към управлението на информацията.

Индустрията е натрупала прекалено много акроними, но всички те описват в различна степен едни и същи процеси и изисквания за данни. Тяхната реална полезност за опростяване на работните процеси остава под въпрос.

Докато понятията и маркетинговите съкращения идват и си отиват, самите процеси на валидиране на изискванията към данните завинаги ще останат неразделна част от бизнес процесите. Вместо да създава все по-специализирани формати и правила, строителният бранш трябва да се обърне към инструменти, които вече са доказали своята ефективност в други области като финансите, промишлеността и ИТ.

Изобилието от термини, съкращения и формати създава илюзията за дълбоко разработени процеси на цифрово конструиране. Маркетинговите понятия и сложната терминология обаче често крият една проста, но неудобна истина: данните остават трудно достъпни, лошо документирани и твърдо обвързани с конкретни софтуерни решения.

За да се излезе от този омагьосан кръг от съкращения и формати заради самите формати, е необходимо да се гледа на CAD (BIM) системите не като на магически инструменти за управление на информацията, а като на това, което те всъщност са - специализирани бази данни. И именно през тази призма може да се разбере къде свършва маркетингът и къде започва истинската работа с информация.



## ГЛАВА 6.4.

# ПАРАМЕТРИЗИРАНЕ НА ДИЗАЙНА И ИЗПОЛЗВАНЕ НА LLM ЗА РАБОТА С CAD

## Илюзията за уникалност на CAD данните (BIM): пътят към анализи и отворени формати

Съвременните CAD (BIM) платформи значително промениха подхода към управлението на информацията за проектиране и строителство. Докато преди тези инструменти се използваха главно за създаване на чертежи и 3D модели, днес те функционират като пълноценни хранилища на данни за проекта. В рамките на концепцията за единен източник на истината параметричният модел все повече се превръща в основен, а често и в единствен източник на информация за проекта, като се гарантира неговата цялост и актуалност през целия жизнен цикъл на проекта.

Основната разлика между CAD - (BIM -) платформите и другите системи за управление на строителни данни е необходимостта от специализирани инструменти и API за достъп до информацията (единственият източник на истината). Тези бази данни не са универсални в традиционния смисъл на думата: вместо отворена структура и гъвкава интеграция, те са затворена среда, твърдо свързана с конкретна платформа и формат.

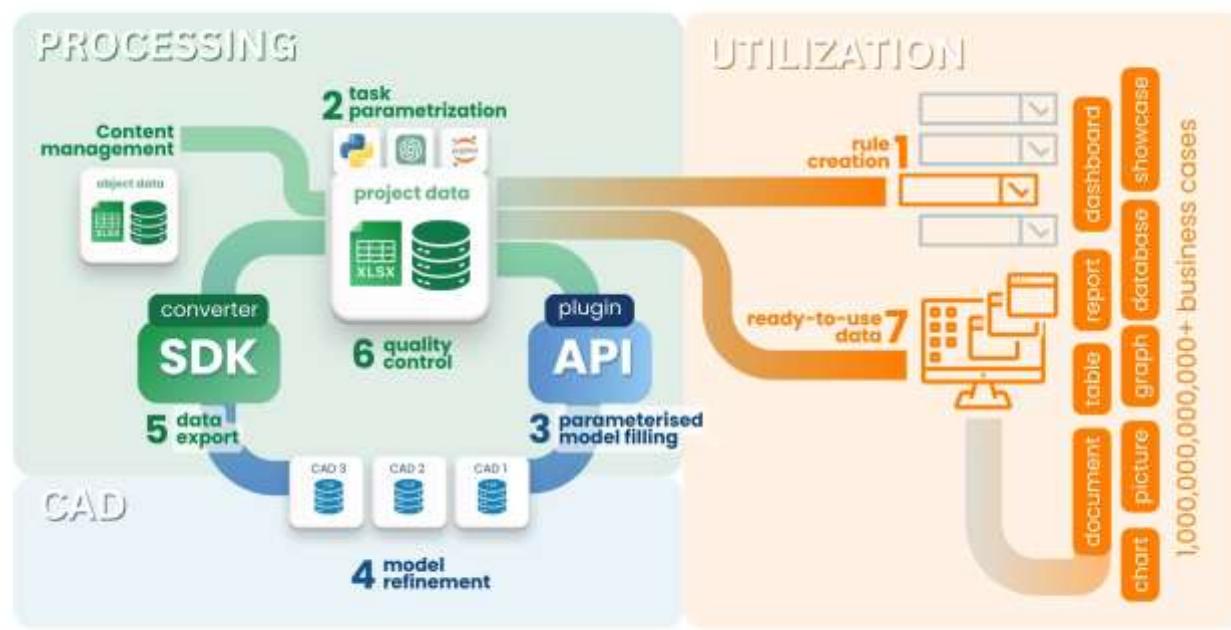
Въпреки сложността на работата с CAD -данни има един по-важен въпрос, който надхвърля техническата реализация: какво всъщност представляват CAD базите данни (BIM)? За да се отговори на този въпрос, е необходимо да се отиде отвъд обичайните съкращения и понятия, наложени от разработчиците на софтуер. Вместо това си струва да се съсредоточим върху същността на работата с информация за проекта: данните и тяхната обработка.

Бизнес процесът в строителството не започва с работа в CAD - или BIM - инструменти, а с формирането на изискванията към проекта и моделирането на данни. Първо се определят параметрите на задачата: списъкът с обекти, техните първоначални характеристики и гранични стойности, които трябва да се вземат предвид при решаването на конкретна задача. Едва след това се създават модели и елементи в CAD (BIM) системите въз основа на зададените параметри

Процесът, който предхожда създаването на информация в базите данни на CAD - (BIM-), е напълно еднакъв с процеса на моделиране на данни, който беше разгледан подробно в четвъртата част на книгата и в главата "Моделиране на данни: концептуален, логически и физически модел" (фиг. 4.3-1).

Както при моделирането на данни създаваме изисквания за данните, които по-късно искаме да обработим в базата данни, така и за CAD базите данни мениджърите създават изисквания за проектиране под формата на няколко колони на таблици или списъци от двойки ключ-стойност (фиг. 6.4-1, стъпки 1-2). И само въз основа на тези първоначални параметри, използвайки API автоматично или ръчно, проектантът създава (или по-скоро усъвършенства) обекти в CAD- (BIM) базите

данни (стъпки 3-4), след което те отново се проверяват за съответствие с първоначалните изисквания (стъпки 5-6). Този процес - дефиниране → създаване → валидиране → коригиране (стъпки 2-6) - се повтаря итеративно, докато качеството на данните, също както при моделирането на данни, достигне желаното ниво за целевата система - документи, таблици или табла (стъпка 7).

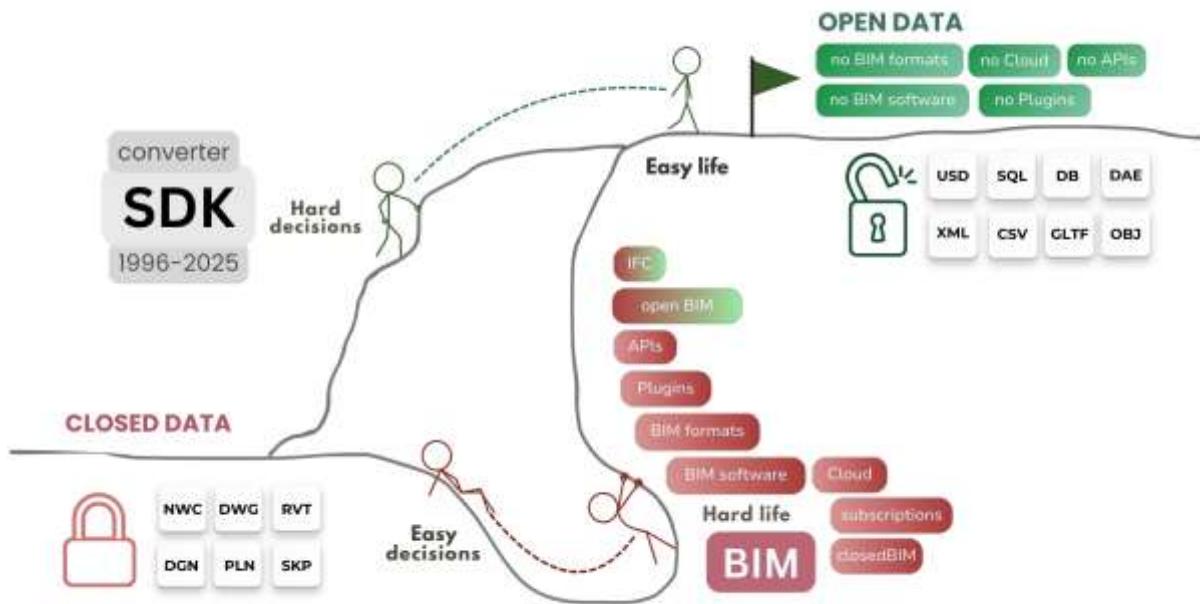


Фиг. 6.4-1 Цикълът на информационно насищане на базите данни за бизнес процесите при изпълнението на строителни проекти.

Ако разглеждаме CAD (BIM) като механизъм за прехвърляне на параметри под формата на набор от двойки ключ-стойност, генерирали въз основа на изисквания, определени извън средата за проектиране (фиг. 6.4-1, стъпки 1-2), фокусът на дискусията се извества от конкретни софтуерни решения и техните ограничения към по-фундаментални аспекти - структура на данните, модели на данните и изисквания към данните. По същество става дума за насищане с параметри на базата данни и за класическия процес на моделиране на данни (стъпки 2-3 и 5-6). Единствената разлика е, че поради затворения характер на CAD-базите данни и особеноностите на използвани формати този процес е съпроводен с използването на специализирани BIM-инструменти. Възниква въпросът: в какво се състои уникалността на BIM, ако в други отрасли няма подобни подходи?

През последните 20 години BIM се позиционира като нещо повече от един източник на данни. Пакетът CAD-BIM често се предлага на пазара като параметричен инструмент с интегрирана по своята същност база данни [64], способен да автоматизира процесите на проектиране, моделиране и управление на жизнения цикъл на строителните проекти. В действителност обаче BIM се е превърнал по-скоро в инструмент за задържане на потребителите на платформата на доставчика, отколкото в удобен метод за управление на данни и процеси.

В резултат на това CAD- (BIM-) данните са изолирани в рамките на техните платформи, като информацията за проекта се крие зад собствени API и геометрични ядра. Това лишава потребителите от възможността за независим достъп до бази данни и извлечане, анализиране, автоматизиране и прехвърляне на данни към други системи, заобикаляйки екосистемите на доставчиците.



Фиг. 6.4-2 В строителството съвременните формати изискват сложни геометрични ядра, ежегодно актуализиран API и специални лицензи за CAD -(BIM-) софтуер.

Фирмите, които работят със съвременни CAD инструменти, трябва да използват същия подход за работа с данни, който самите доставчици на CAD използват на практика без изключение: преобразуване на данни с помощта на SDK - инструменти за обратно инженерство, срещу които CAD производителите се борят от 1995 г. [75]. Имайки пълен достъп до CAD базата данни и използвайки инструментите за обратно инженерство, можем да получим [127] плосък набор от същности с атрибути и да ги експортираме във всеки удобен отворен формат (фиг. 6.4-2), включващ както геометрията, така и параметрите на елементите на конструкцията. Този подход коренно променя парадигмата на работа с информация - от файлово ориентирана към ориентирана към данните архитектура:

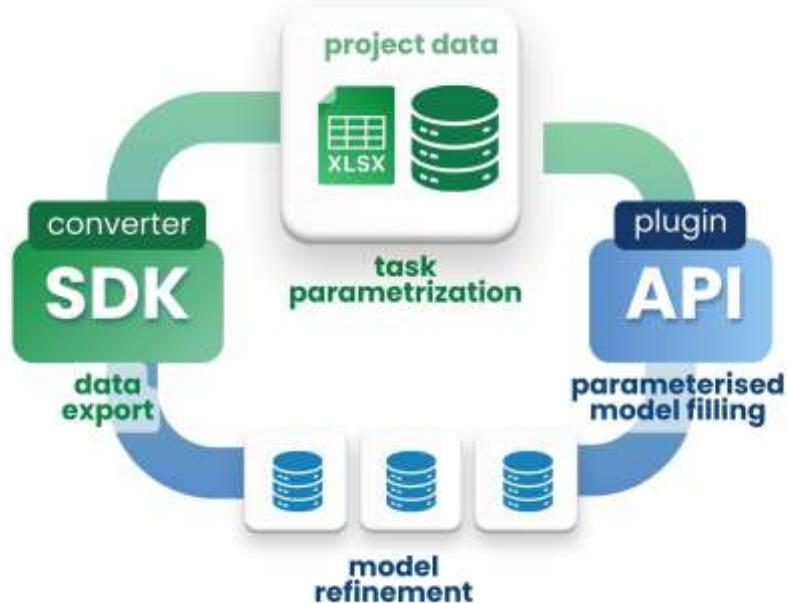
- Форматите за данни, като RVT, IFC, PLN, DB1, CP2, CPIXML, USD, SQLite, XLSX, PARQUET и други, съдържат идентична информация за елементите на един и същ проект. Това означава, че познаването на конкретен формат и неговата схема не трябва да бъде пречка за работа със самите данни.
- Данните от всякакви формати могат да се комбинират в една отворена структурирана и гранулирана структура (фигура 9.1-10), съдържаща триъгълната геометрия на MESH и свойствата на всички обектни същности, без ограниченията на геометричните ядра.
- Анализът на данни се стреми към универсалност: използвайки отворени данни, можете да работите с данни за проекта, независимо от използвания формат.
- Минимизиране на зависимостта от API и приставки на доставчиците: работата с данни вече не зависи от уменията за работа с API.

Когато и CAD -изискванията за данни се трансформират в лесни за анализиране структурирани

формати за представяне - разработчиците вече не зависят от специфични схеми за данни и затворени екосистеми.

## Проектиране чрез параметри: бъдещето на CAD и BIM

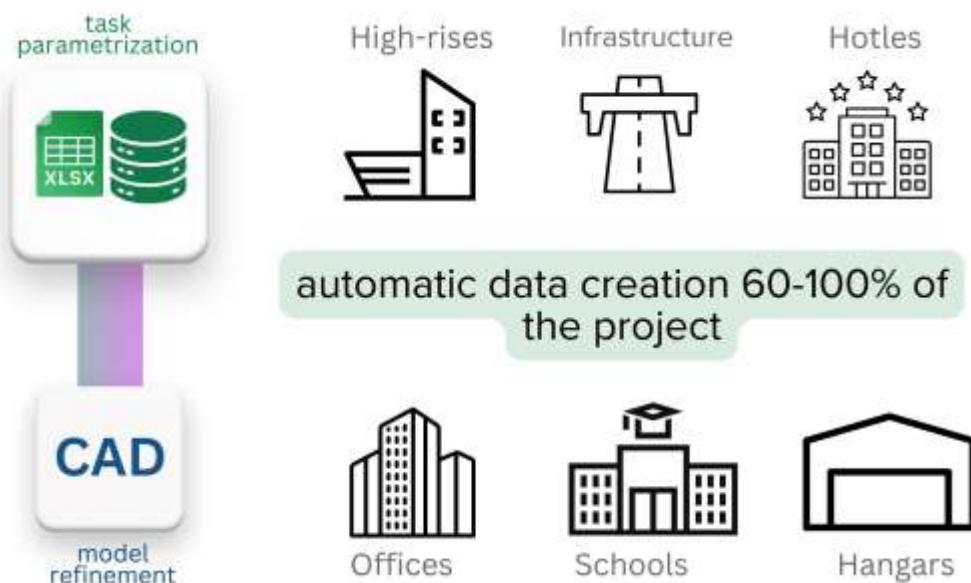
Нито един строителен проект в света не е започнал с програма CAD. Преди чертежът или моделът да придобие форма в CAD, той преминава през етапа на концептуализация (фиг. 6.4-1, етапи 1-2), където вниманието е насочено към параметрите, които определят основната идея и логика на бъдещия обект. Този етап съответства на концептуалното ниво при моделирането на данни (фиг. 4.3-6). Параметрите могат да съществуват единствено в съзнанието на дизайнера, но в идеалния случай те са подредени под формата на структурирани списъци, таблици или се съхраняват в бази данни (фиг. 6.4-3), което позволява прозрачност, възпроизвежданост и по-нататъшна автоматизация на процеса на проектиране.



**Фиг. 6.4-3** Процесът на проектиране е итеративен процес на попълване на CAD базата данни с информация отвън, като се използват изискванията във веригата на стойността.

Преди да се започне самото CAD моделиране (логическа и физическа фаза на моделиране на данни (фиг. 4.3-7)), е важно да се определят граничните параметри, които служат за основа на проекта. Тези параметри, както и другите изисквания, се събират от самия край на веригата за използване на данни (напр. системите) и чрез тях вече се определят ограниченията, целите и ключовите характеристики на бъдещите обекти в проекта.

Самото моделиране може да бъде напълно автоматизирано с 60-100% с помощта на инструменти за параметрично моделиране (фиг. 6.4-3), ако изискванията са добре дефинирани. Веднага след като проектът бъде описан под формата на параметри, неговото формиране става технически осъществимо, например с помощта на визуални езици за програмиране като Grasshopper Dynamo, вградени в съвременни CAD - среди или безплатни решения в Blender, UE, Omniverse.



Фиг. 6.4-4 По-голямата част от типизираните проекти днес вече се създават напълно автоматично благодарение на инструментите за параметрично програмиране.

Днес големите индустриални и типови проекти вече не се създават от ръцете на дизайнерския отдел, а чрез параметрични инструменти и визуално програмиране. Това дава възможност да се изгради модел въз основа на данни, а не на субективните решения на конкретен дизайнер или менеджър.

*Съдържанието предхожда дизайна. Дизайн без съдържание не е дизайн, а декорация [128].*

- Джефри Зелдман, уеб дизайнер и предприемач

Процесът не започва с изготвяне на чертеж или 3D -моделиране, а с формиране на изискванията. Именно изискванията определят какви елементи ще се използват в проекта, какви данни трябва да се прехвърлят към други отдели и системи. Само наличието на структурирани изисквания дава възможност за автоматична проверка на моделите на регулярна основа (например дори на всеки 10 минути, без да се отвлича вниманието на проектанта от работата му).

Може би в бъдеще CAD- (BIM-) системата ще се превърне само в интерфейс за попълване на базата данни и няма да има значение в кой CAD инструмент извършва моделирането (физическо ниво).

По същия начин в машиностроенето често се използва 3D моделиране, но то не е необходим или задължителен елемент от проекта. В повечето случаи класическата 2D документация е достатъчна и необходимият информационен модел се създава въз основа на нея. Този модел се сглобява от

компоненти, структурирани съгласно индустритлните стандарти, и съдържа цялата необходима информация за разбиране на организацията на проектирането и производството. След това информационният модел на завода се използва за създаване на фабричен информационен модел, към който се добавят специфични продукти и блок-схеми, които вече са ориентирани към нуждите на технолозите. Целият процес може да бъде организиран без излишна сложност, без да се претоварва системата с 3D графики там, където те не дават реални предимства.

Важно е да се разбере, че самият 3D модел и CAD -системата не трябва да играят главната роля - те са само инструмент за количествен и геометричен анализ. Всички други параметри, с изключение на геометрията, които описват обекта, трябва да се съхраняват и обработват извън CAD средата, ако е възможно (BIM).

Проектирането чрез параметри не е просто тенденция, а неизбежното бъдеще на строителната индустрия. Вместо да създават ръчно сложни 3D -модели, проектантите ще работят с данни, ще ги валидират и ще автоматизират процесите, доближавайки строителството до света на програмирането. С течение на времето процесите на проектиране ще бъдат изградени на принципите на софтуерната разработка:

- Създаване на изисквания → Създаване на модел → Качване на сървъра → Утвърждаване на промени → Заявка за изтегляне
- Заявките за изтегляне автоматично извършват проверки на модела спрямо изискванията, които са били създадени преди или по време на проектирането като част от заявката за изтегляне.
- След проверка на качеството на данните и одобрение, промените се въвеждат в проекта, в общата база данни или се прехвърлят автоматично в други системи.

Още сега в машиностроенето подобни промени в дизайна започват със създаването на известие за промяна. Подобна схема очаква и строителната индустрия: проектирането ще бъде итеративен процес, в който всяка стъпка се подкрепя от параметрични изисквания. Такава система ще позволи на проектантите да създават автоматични проверки и автоматични заявки за изтегляне на специфични изисквания.

Дизайнърът на бъдещето е преди всичко оператор на данни, а не ръчен моделиер. Неговата задача е да изпълни проекта с параметрични обекти, при които геометрията е само един от атрибути.

Именно разбирането на значението на моделирането на данни, класификацията и стандартизирането, които бяха разгледани подробно в предишните глави на книгата, ще изиграе важна роля в трансформацията. Регламентите за проектиране на бъдещето ще бъдат формализирани като двойки параметри ключ-стойност под формата на XLSX или XML -схеми.

Бъдещето на строителната индустрия е свързано със събирането на данни, анализирането им, валидирането им и автоматизирането на процесите с помощта на инструменти за анализ. BIM (или CAD) не е крайната цел, а само етап от еволюцията. Когато специалистите осъзнават, че могат да работят директно с данни, заобикаляйки традиционните CAD инструменти, самият термин "BIM" постепенно ще отстъпи място на концепциите за използване на структурирани и детализирани данни за строителните проекти.

Един от ключовите фактори, ускоряващи трансформацията, е появата на големи езикови модели (LLM) и на базираните на тях инструменти. Тези технологии променят начина, по който се обработват данните за проектиране, като дават възможност за достъп до информация, без да е необходимо задълбочено познаване на API или решенията на доставчиците. С LLM процесът на създаване на изисквания и взаимодействие с CAD данни става интуитивен и достъпен.

## **Възникване на LLM в процесите на обработка на данни в CAD проектирането**

В допълнение към разработването на инструменти за достъп до CAD бази данни и отворени и опростени CAD -формати, появата на LLM -инструменти (Large Language Models) революционизира обработката на данни за проектиране. Докато в миналото достъпът до информация се осъществяваше главно чрез сложни интерфейси и изискваше умения за програмиране и познаване на API, сега е възможно да се взаимодейства с данните с помощта на естествен език.

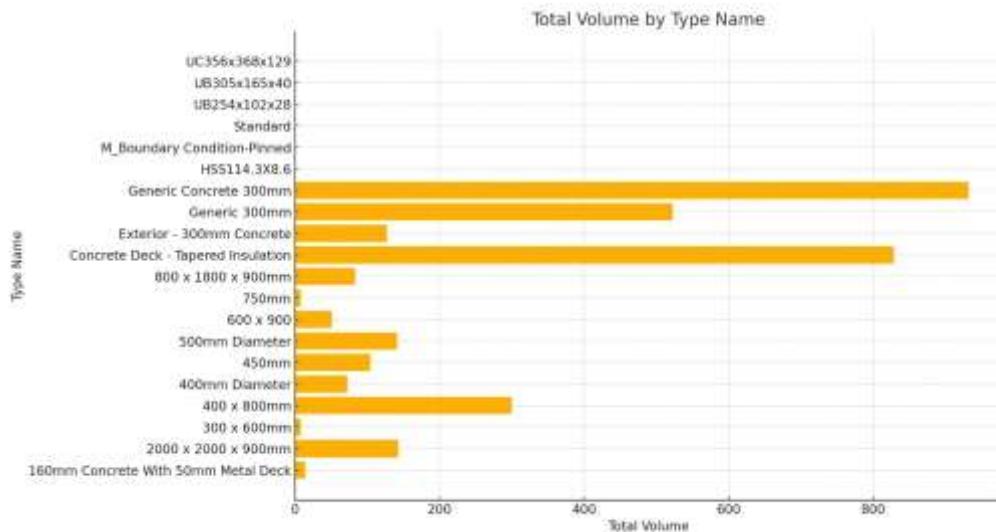
Инженерите, мениджърите и проектантите, които нямат техническо образование, могат да получат необходимата информация от данните за проекта, като формулират заявки на обикновен език. При условие че данните са структурирани и достъпни (фиг. 4.1-13), достатъчно е да зададете въпрос в LLM чат като: "Покажете в таблица с групиране по тип всички стени с обем над 10 куб. м" - и моделът автоматично ще преобразува това запитване в SQL или код в Pandas, генерирайки обобщена таблица, графика или дори готов документ.

По-долу са дадени някои реални примери за това как LLM -моделите взаимодействват с проектни данни, представени в различни CAD- (BIM-) формати.

- ❷ Пример за заявка в LLM чат към CAD проект във формат RVT след преобразуване (фиг. 4.1-13) в таблична рамка от данни (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN или друга):

Групирайте данните в рамката на данните, получена от файла RVT, по "Type name" (Име на типа) при сумиране на параметъра "Volume" (Обем) и покажете броя на елементите в групата. И моля, покажете всичко това като хоризонтална хистограма без нулеви стойности.

- 💡 Отговор на LLM като хоризонтална стълбовидна графика (формат PNG):



Фигура 6.4-5 Вместо 17 щраквания с мишката или 40 реда код чрез използване на приставки, в LLM незабавно извлечаме таблицата QTO чрез текстова заявка.

- 💡 За да генерираме таблица QTO на видовете стени с обща площ и количество от категорията "Стени", нека формулираме текстово запитване за LLM -chat:

Вземете само тези елементи от рамката с данни за проекта, които имат "OST\_Walls" в параметъра "Category" (Категория), групирайте ги по "Type Name" (Наименование на типа), съберете стойността на колоната "Area" (Площ), добавете количеството и ги покажете в таблицата, като премахнете нулевите стойности.

- ☒ Отговор на LLM под формата на завършена таблица за QTO:

Type Name	Total Area	Count
CL_W1	393.12 sq m	10
Cavity wall sliders	9.37 sq m	1
Foundation - 300mm Concrete	30.90 sq m	1
Interior - 165 Partition (1-hr)	17.25 sq m	3
Interior - Partition	186.54 sq m	14
Retaining - 300mm Concrete	195.79 sq m	10
SH_Curtain wall	159.42 sq m	9
SIP 202mm Wall - conc clad	114.76 sq m	4
Wall - Timber Clad	162.91 sq m	8

**Фиг. 6.4-6 Създаването на таблица QTO на естествен език осигурява същото качество на резултата, както при използването на CAD - (BIM- ) инструменти.**

- ☒ Нека да направим заявка за проекта във формат IFC след преобразуване в таблична рамка от данни и да въведем подобна текстова заявка във всеки чат на LLM:

Вземете само тези елементи от проекта, които имат стойности на ниво 1 и ниво 2 в параметъра "Родител", и вземете елементите, които имат стойности на IfcSlab в параметъра "Категория", след това групирайте тези елементи по параметъра "Тип на обекта", съберете стойностите в параметъра "PSet\_RVT \_Размерна област" и ги покажете като кръгова диаграма.

- ❑ Отговор на LLM като завършена кръгова диаграма на групите елементи от данните на IFC :

For Floor:127mm Slab on Grade, Floor:150mm Exterior Slab on Grade,  
 Floor:Finish Floor - Ceramic Tile, Floor:Finish Floor - Wood, and  
 Floor:Residential - Wood Joist with Subflooring



**Фигура 6.4-7 Резултатът от заявка за IFC данни в структуриран формат може да бъде всеки тип графика, който е удобен за разбиране на данните.**

Зад всяко от получените готови решения (фиг. 6.4-5 - фиг. 6.4-7) стоят десетина реда код в Python, използващ библиотеката Pandas. Полученият код може да се копира от чата на LLM и да се използва във всеки локален или онлайн IDE, за да се получат идентични резултати извън чата на LLM.

В същия чат на LLM можем да работим не само с проекти, получени от 3D CAD (BIM) формати, но и с плоски чертежи във формат DWG, към които можем да отправим запитване в чата на LLM, за да покажем например данни за групи елементи под формата на линии или 3D геометрии след преобразуване в структуриран вид.

## Автоматизиран анализ на DWG -файлове с LLM и Pandas

Процесът на обработка на данни от DWG -файлове поради неструктурания характер на информацията - винаги е бил сложна задача, изискваща специализиран софтуер и често ръчен анализ. Въпреки това, с развитието на изкуствения интелект и инструментите за LLM, стана възможно да се автоматизират много стъпки, този, днес, предимно ръчен процес. Нека разгледаме една реална тръбопроводна линия на заявки към LLM (в този пример ChatGPT) за работа с DWG чертежи, които позволяват да се работи с проекта:

- ❑ Филтриране на DWG данни по слой, ID и координати
- ❑ Визуализиране на геометрията на елементите
- ❑ Автоматично анотиране на чертежи въз основа на параметри

- Разширяване на полилинииите на стените до хоризонталната равнина
- Създаване на интерактивни 3D - визуализации на равнинни данни
- Структуриране и анализиране на строителни данни без сложни CAD -инструменти

В нашия случай процесът на изграждане на тръбопровода започва с последователно генериране на код чрез LLM. Първо се генерира заявка, описваща задачата. ChatGPT генерира Python -код, който се изпълнява и анализира, като резултатът се показва в чат стаята. Ако резултатът не е според очакванията, заявката се коригира и процесът се повтаря

**Конвойерът** е последователност от автоматизирани стъпки, извършвани за обработка и анализ на данни. В такъв процес всяка стъпка приема данни като вход, извършва трансформации и предава резултата на следващата стъпка.

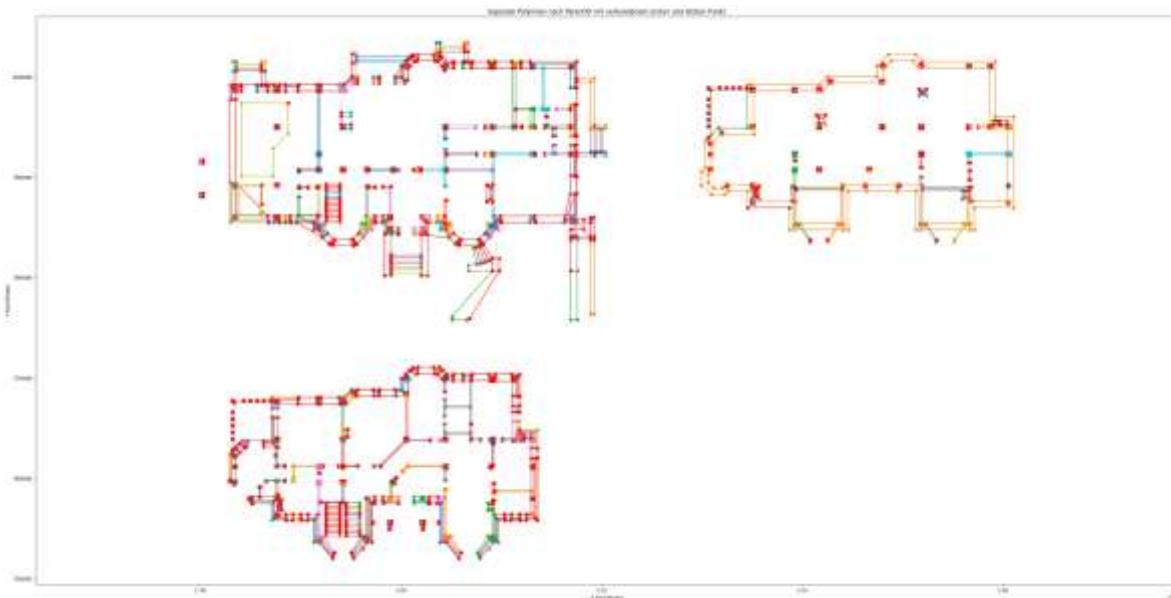
След получаване на желания резултат кодът се копира от LLM и се вмъква в кода под формата на блокове в някоя от удобните IDE, в нашия случай в платформата Kaggle.com. Получените фрагменти от кода се обединяват в единен конвойер, който автоматизира целия процес - от зареждането на данните до окончателния им анализ. Този подход позволява бързо разработване и мащабиране на аналитични процеси без задълбочени познания по програмиране. Пълният код на всички фрагменти по-долу, заедно с примерни заявки, може да бъде намерен в платформата Kaggle.com чрез търсене на "DWG Analyse with ChatGPT | DataDrivenConstruction" [129].

Нека да започнем процеса на работа с DWG данни, след преобразуването им в структуриран вид (фиг. 4.1-13), с класическа стъпка - групиране и филтриране на всички чертожни данни, необходими за нашата задача елементи на стената, по-конкретно полилинии (параметърът "ParentID" позволява групиране на линиите в групи), които в параметъра (колона на рамката на данните) "Layer" имат стойност на низ, съдържащ следната комбинация от букви (RegEx) - "wall".

- ❷ За да получите кода за подобна задача и резултата под формата на снимка, трябва да напишете следната заявка в LLM:

Първо, проверете дали рамката с данни, получена от DWG, съдържа определените колони: "Layer", "ID", "ParentID" и "Point". След това филтрирайте идентификаторите от колоната "Layer", които съдържат символа "wall". Намерете елементите в колоната "ParentID", които отговарят на тези идентификатори. Дефинирайте функция за почистване и разделяне на данните в колоната "Point". Това включва премахване на скобите и разделяне на стойностите на координати "x", "y" и "z". Изведете данните на графиката, като използвате matplotlib. За всеки уникален "ParentID" начертайте отделна полилиния, свързваща координатите на "Point". Уверете се, че първата и последната точка са свързани, ако е възможно. Задайте подходящи етикети и заглавия, като се уверите, че осите x и y са еднакво мащабирани.

- ❸ Отговорът LLM ще ви даде готова картина, зад която се крие кодът на Python, който я е генерира:



Фиг. 6.4-8 Кодът LLM извличе всички линии на слоя "стена" от файла DWG, изчисти координатите им и построи полилините, като използва една от библиотеките на Python.

- Сега нека добавим към линиите параметъра за площ, който всяка полилиния има в своите свойства (в една от колоните на рамката с данни):

Сега вземете само един "ParentID" от всяка полилиния - намерете този идентификатор в колоната "ID", вземете стойността на "Area", разделете я на 1 000 000 и добавете тази стойност към графиката.

- Отговорът на LLM ще покаже нова графика, в която всяка полилиния ще има надпис с нейната площ:

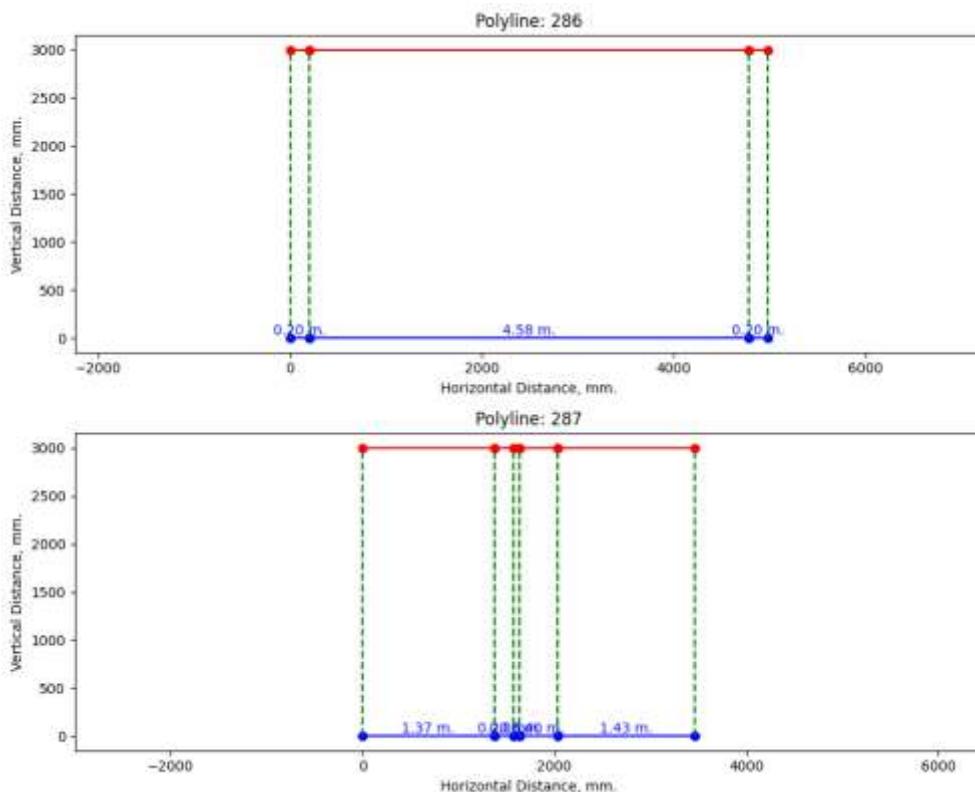


**Фиг. 6.4-9 LLM е добавил код, който взема стойностите на площта за всяка полилиния и я добавя към изображението с визуализация на линиите.**

- След това ще превърнем всяка полилиния в хоризонтална линия, ще добавим успоредна линия на височина 3000 mm и ще ги свържем в една равнина, за да покажем по този начин разположението на повърхностите на стените елементи:

Трябва да вземете всички елементи от колоната "Layer" със стойност "wall". Вземете тези идентификатори като списък от колоната "ID" и намерете тези идентификатори от цялата рамка от данни в колоната "ParentID". Всички елементи са линии, които са обединени в една полилиния. Всяка линия има различна геометрия x, у на първата точка в колоната "Point" (Точка). Трябва да вземете последователно всяка полилиния и от точката 0,0 хоризонтално да начертаете дължината на всеки сегмент от полилинията. дължината на всеки сегмент от полилинията в една линия. След това начертайте абсолютно същите линии, само че с 3000 по-високи, като свържете всички точки в една равнина.

- Отговорът на LLM ще изведе код, който ви позволява да чертаете чертежи на стени в равнината:

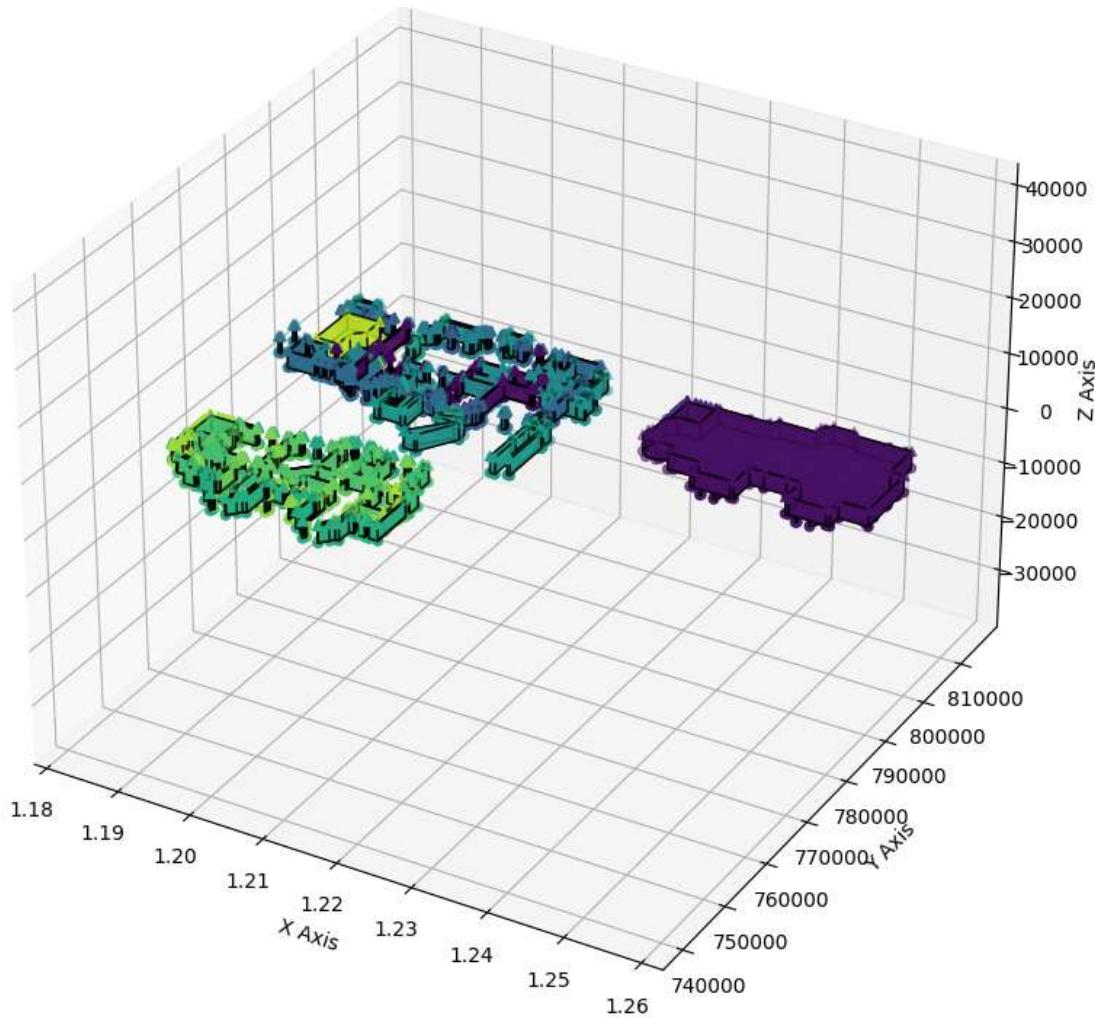


**Фиг. 6.4-10** Превръщаме всяка полилиния с помощта на подсказки в макет, който визуализира плоскостите на стените директно в чата на LLM.

- Сега нека преминем от 2D проекция към 3D - моделиране на стени от плоски линии чрез свързване на горния и долния слой полилиниии:

Визуализирайте елементите на стената в 3D, като свържете полилиниии на височини  $z = 0$  и  $z = 3000 \text{ mm}$ . За да създадете затворена геометрия, представляваща стените на сградата. Използвайте инструмента за 3D графики Matplotlib.

- LLM ще генерира интерактивен 3D -граф, в който всяка полилиния ще бъде представена като набор от равнини. Потребителят ще може да се движи свободно между елементите с помощта на компютърна мишка, като изследва модела в 3D режим чрез копиране на кода от чата в IDE:



**Фиг. 6.4-11** LLM помогна за създаването на код [129] за визуализиране на плоски чертожни линии в 3D изглед, който може да се изследва в 3D визуализатора в рамките на IDE.

За да се изгради логична и възпроизвеждаема тръбопроводна линия - от първоначалното преобразуване и зареждане на DWG -файла до крайния резултат - се препоръчва след всяка стъпка да се копира генеририаният LLM -блок код в IDE. По този начин не само проверявате резултата в чата, но и го стартирате веднага във вашата среда за разработка. Това ви позволява да изграждате процеса последователно, като го отстранявате и адаптирате при необходимост.

Можете да намерите пълния код на тръбопровода на всички фрагменти (фигури 6.4-8 до фигури 6.4-11) заедно с примерни заявки на платформата Kaggle.com, като потърсите "DWG Analyse with ChatGPT | DataDrivenConstruction". [129]. На Kaggle можете не само да разглеждате кода и използвате подсказки, но и да копирате и тествате бесплатно цялата Pipeline с оригиналните DWG

dataframes в облака, без да се налага да инсталирате допълнителен софтуер или самата IDE.

Подходът, представен в тази глава, ви позволява напълно да автоматизирате проверката, обработка и генерирането на документи въз основа на DWG -проекти. Разработеният конвейер е подходящ както за обработка на отделни чертежи, така и за пакетна обработка на десетки, стотици и хиляди DWG-файлове с автоматично генериране на необходимите отчети и визуализации за всеки проект.

Процесът може да бъде организиран по последователен и прозрачен начин: първо данните от CAD -файла се преобразуват автоматично във формат XLSX, след което се зареждат в рамка от данни, последвано от групиране, проверка и генериране на резултати - всичко това се реализира в една тетрадка на Jupyter или Python -скрипт, във всяка популярна IDE. Ако е необходимо, процесът може лесно да бъде разширен чрез интегриране със системи за управление на проектната документация: CAD файловете могат да бъдат автоматично извлечени в съответствие с определени критерии, резултатите могат да бъдат връщани обратно в системата за съхранение, а потребителите могат да бъдат уведомявани, когато резултатите са готови - чрез електронна поща или месинджъри.

Използването на чатове и агенти на LLM за работа с проектни данни намалява зависимостта от специализирани CAD -програми и ви позволява да извършвате анализ и визуализация на архитектурни проекти без необходимост от ръчно взаимодействие с интерфейса - без кликвания с мишката и запомняне на сложна навигация в менюто.

С всеки изминал ден строителната индустрия ще чува все повече и повече за LLM, гранулирани структурирани данни, DataFrames и колонарни бази данни. Унифицираните двуизмерни DataFrames, формирани от различни бази данни и CAD формати, ще бъдат идеалното гориво за съвременните аналитични инструменти, с които активно боравят специалистите в други отрасли.

Самият процес на автоматизация ще бъде значително опростен - вместо да изучавате API на затворени нишови продукти и да пишете сложни скриптове за анализ или преобразуване на параметри, сега ще е достатъчно да формулирате задачата под формата на набор от отделни текстови команди, които ще бъдат сложени в необходимия Pipeline или Workflow-процес за необходимия език за програмиране, който работи бесплатно на почти всяко устройство. Вече няма да се чакат нови продукти, формати, плъгини или актуализации от производителите на CAD- (BIM-) инструменти. Инженерите и строителите ще имат възможност да работят самостоятелно с данни, използвайки прости, бесплатни и лесни за разбиране инструменти, подпомагани от LLM чатове и агенти.

## Следващи стъпки: преминаване от затворени формати към отворени данни

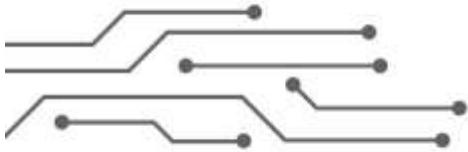
При работа с данните за проектиране на бъдещето едва ли някой ще трябва да разбира геометричните ядра на патентовани инструменти или да изучава стотици несъвместими формати, съдържащи една и съща информация. Въпреки това, без да се разбере защо преходит към отворени структурирани данни е важен, е трудно да се аргументира използването на нови бесплатни инструменти, отворени данни и подходи, които е малко вероятно да бъдат наследчавани от производителите на софтуер.

В тази глава обсъдихме основните характеристики на данните от CAD (BIM), техните ограничения и възможности, както и това, че въпреки маркетинговите обещания на доставчиците, инженерите и проектантите се сблъскват с трудности при извлечането, прехвърлянето и анализирането на проектната информация всеки ден. Разбирането на архитектурата на тези системи и запознаването с алтернативни подходи - базирани на отворени формати и автоматизация чрез LLM - може да улесни значително живота дори на един специалист, да не говорим за компаниите. За да обобщим тази част, си струва да подчертаем основните практически стъпки, които ще ви помогнат да приложите подходите, разгледани в този раздел, в ежедневните си задачи:

- Разширете набора си от инструменти за работа с данни от проекти
  - Разгледайте наличните приставки и помощни програми за извлечане на данни от използваните от вас CAD - (BIM-) системи.
  - Разгледайте наличните SDK и API, които ви позволяват да автоматизирате извлечането на данни от затворени формати, без да се налага да отваряте ръчно специализиран софтуер.
  - Усвояване на основни умения за работа с отворени непараметрични геометрични формати (OBJ, glTF, USD, DAE) и съответните библиотеки с отворен код
  - Опитайте се да измислите система за съхраняване на метаданни за проекта, отделно от геометрията, извън CAD (BIM) решения, за да опростите анализа и интеграцията с други системи.
  - Използвайте LLM, за да автоматизирате преобразуването на данни между формати.
- Създаване на собствени процеси за обработка на информация за проекта
  - Започнете да описвате задачи и изисквания за моделиране чрез параметри и техните стойности в прости и структурирани формати
  - Създаване на лична библиотека от скриптове или блокове код за често извършвани операции
- Насърчаване на използването на отворени стандарти в работата ви
  - Поканете колеги и партньори да споделят данни в отворени формати, които не са ограничени от екосистемата на доставчика на софтуер.
  - Демонстриране на ползите от използването на структурирани данни с конкретни примери
  - Иницииране на дискусии относно проблемите със затворените формати и възможните решения

Дори и да не можете да промените политиката на вашата компания по отношение на CAD - (BIM-) платформите, личното разбиране на принципите на работа с проектни данни в отворени формати ще ви позволи значително да повишите ефективността на работата си. Създавайки свои собствени инструменти и методи за извлечане и трансформиране на данни от различни формати, вие не само

оптимизирате работните си процеси, но и придобивате гъвкавост, за да заобиколите ограниченията на стандартните софтуерни решения.



## VII ЧАСТ

### **ВЗЕМАНЕ НА РЕШЕНИЯ, БАЗИРАНИ НА ДАННИ, АНАЛИЗ, АВТОМАТИЗАЦИЯ И МАШИННО ОБУЧЕНИЕ.**

Седмата част е посветена на анализа на данни и автоматизацията на процесите в строителната индустрия. В нея се обсъжда как данните стават основа за вземане на решения и се обясняват принципите на визуализиране на информацията за ефективен анализ. Погледнато са описани ключовите показатели за ефективност (KPI), методите за оценка на възвръщаемостта на инвестициите (ROI) и създаването на информационни табла за наблюдение на проекти. Специално внимание е отделено на процесите ETL (Extract, Transform, Load) и тяхното автоматизиране с помощта на тръбопроводи (Pipeline) за превръщане на разнородни данни в структурирана информация за анализ. Обсъждат се инструменти за оркестрация на работни потоци, като Apache Airflow, Apache NiFi и n8n, които позволяват изграждането на автоматизирани потоци от данни без задълбочени познания по програмиране. Значителна роля играят големите езикови модели (Large Language Models - LLMs) и тяхното използване за опростяване на анализа на данни и автоматизиране на рутинни задачи.

## ГЛАВА 7.1.

### АНАЛИЗ НА ДАННИ И ВЗЕМАНЕ НА РЕШЕНИЯ, ОСНОВАНИ НА ДАННИ

След етапите на събиране, структуриране, почистване и проверка на информацията се получава съгласуван и анализиран набор от данни. Предишните части на книгата обхващат систематизирането и структурирането на разнородни източници - от PDF документи и текстови записи на срещи до CAD модели и геометрични данни. Подробно е описан процесът на проверка и привеждане на информацията в съответствие с изискванията на различни системи и класификатори, отстраняване на дублирания и несъответствия.

Всички изчисления, които се извършват върху тези данни (трета и четвърта част на книгата) - от прости преобразувания до изчисления на времето, разходите и показателите на ESG (пета част) - са обобщени аналитични задачи. Те формират основата за разбиране на текущото състояние на даден проект, оценка на неговите параметри и последващо вземане на решения. В резултат на това данните, в резултат на изчисленията, се превръщат от набор от разнородни записи в управляем ресурс, способен да отговори на ключови бизнес въпроси.

В предишните глави бяха описани подробно процесите на събиране на данни и контрол на качеството за използване в типични бизнес случаи и процеси, специфични за строителната индустрия. Анализът в този контекст е подобен в много отношения на приложенията в други индустрии, но има редица специфични характеристики.

В следващите глави ще бъде описан подробно процесът на анализ на данни, включително етапите на автоматизация от първоначалното получаване на информация и нейното преобразуване до последващото ѝ прехвърляне към целевите системи и документи. Първо ще бъде представена теоретична част, която се фокусира върху избрани аспекти на вземането на решения, основани на данни. След това, в следващите глави, ще бъде представена практическата част, свързана с автоматизацията и изграждането на ETL -Pipeline.

#### **Данните като ресурс при вземането на решения**

Вземането на решения, базирани на данни, често е итеративен процес и започва със систематично събиране на информация от различни източници. Подобно на естествения цикъл, отделните елементи на данните и целите информационни системи постепенно попадат в почвата - натрупват се в информационните хранилища на компаниите (фигура 1.3-2). С течение на времето тези данни, подобно на падналите листа и клони, се трансформират в ценен материал. Мицелът на инженерите и анализаторите на данни организира и подготвя информацията за бъдеща употреба и превръща падналите данни и системи в ценен компост, за да израснат нови филизи и нови системи (фигура 1.2-5).

Тенденциите в широкото използване на анализите в различни индустрии, бележи началото на нова ера, в която работата с данни се превръща в основа на професионалната дейност (фиг. 7.1-1). За специалистите в строителния бранш е важно да се адаптират към тези промени и да бъдат готови да преминат в нова ера - ерата на данните и анализите

Ръчното преместване на данни между таблици и ръчното извършване на изчисления постепенно остават в миналото, отстъпвайки място на автоматизацията, анализа на потока от данни, анализа и машинното обучение. Тези инструменти се превръщат в ключови елементи на съвременните системи за подпомагане на вземането на решения.

В книгата на McKinsey "Рестартиране. Ръководство за преодоляване на конкуренцията в ерата на цифровите технологии и изкуствения интелект" [130] се цитира проучване сред 1330 висши ръководители от различни региони, индустрии и функционални области, проведено през 2022 г. [130], се позовава на проучване, проведено през 2022 г. с участето на 1330 висши ръководители от различни региони, индустрии и функционални области. Според резултатите от него 70% от ръководителите използват усъвършенствани анализи, за да генерират собствени идеи, а 50% прилагат изкуствен интелект за подобряване и автоматизиране на процесите на вземане на решения.



**Фиг. 7.1-1 Анализът на данни и анализът е основният инструмент за увеличаване на скоростта на вземане на решения в една компания.**

Анализът на данни, подобно на разпространението на мицела, прониква в хумуса на минали решения, като помага за свързването на отделните системи и насочва ръководителите към ценни прозрения. Тези знания, подобно на хранителните вещества от изгнилите дървета на системите за данни, подхранват новите решения в компанията, което води до ефективна промяна и качествен информационен растеж, подобно на новите издънки и кълнове, излизящи от богата и здрава почва (фигура 1.2-5).

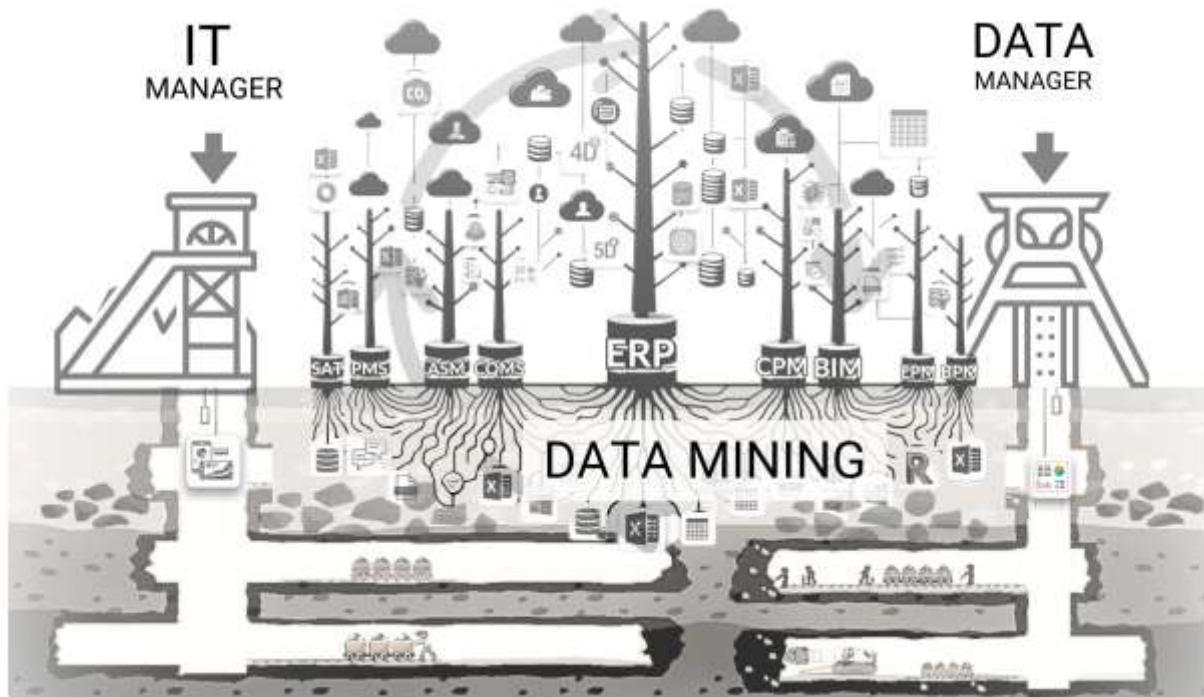
*Числата имат важна история за разказване. Те разчитат на вас да им дадете ясен и убедителен глас [131].*

- Стивън Фю, експерт по визуализация на данни

В средните и малките предприятия работата по извлечане и подготовка на информацията за по-нататъшен анализ днес е изключително трудоемък процес (фигура 7.1-2), сравним с добива на въглища през XVIII век. До неотдавна работата по извлечане и подготовка на данни беше по-скоро запазена за авантюристи, работещи в тясно специализирана ниша с малък и ограничен набор от

инструменти за работа с различни видове данни от неструктурирани, слабо структурирани, смесени и затворени източници.

Лицата, вземащи решения, и мениджърите често нямат опит в работата с разнородни данни и системи, но трябва да вземат решения въз основа на тях. В резултат на това през последните десетилетия вземането на решения, базирани на данни, в съвременната строителна индустрия не прилича толкова на автоматизиран процес, колкото на многодневния ръчен труд на миньора в ранните въглищни мини.



Фиг. 7.1-2 В процеса на извличане на данни експертите преминават през сложен път на подготовка на данните - от почистване до структуриране за последващ анализ.

Макар че съвременните методи за извличане на данни в строителния бранш със сигурност са по-напреднали от примитивните техники на миньорите от XII век, това все още е сложна и високорискова задача, изискваща значителни ресурси и експертни познания, които могат да си позволяят само големите компании. Процесите на извличане и анализ на данни от натрупаното наследство от минали проекти доскоро се извършваха предимно от големи, технологично напреднали компании, които събират и съхраняват данни последователно от десетилетия

Преди това водеща роля в областта на анализите играеха технологично развитите компании, които са трупали данни в продължение на десетилетия. Днес ситуацията се променя: достъпът до данни и инструменти за обработка на данни се демократизира - преди сложните решения вече са достъпни за всички бесплатно.

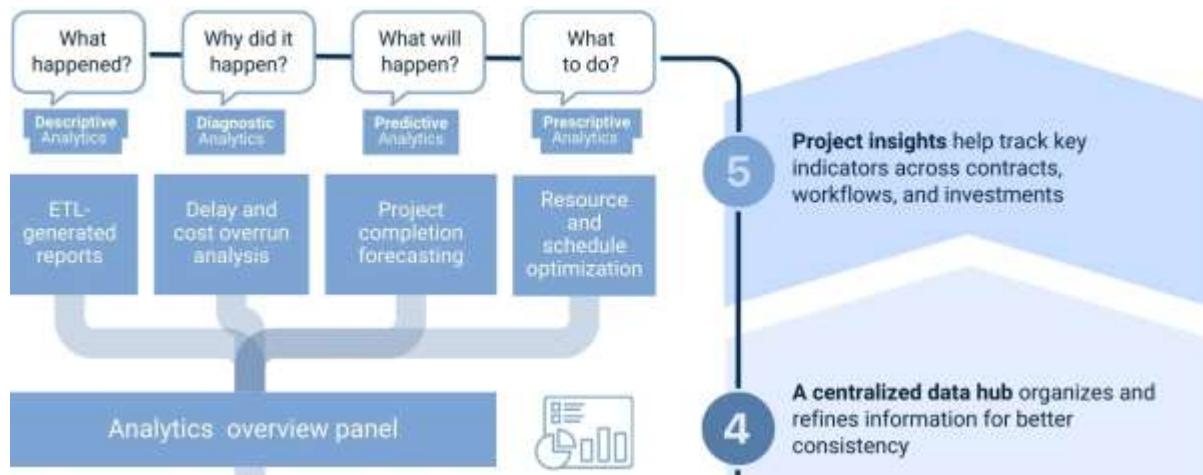
Прилагането на анализи позволява на компаниите да вземат по-точни и информирани решения в реално време. Следващият казус илюстрира как историческите данни могат да помогнат за вземането на финансово обосновани решения:

- ❑ **Ръководител на проекта** - "Сега средната цена на бетона в града е 82€ /m<sup>3</sup>, а в оценката имаме 95 €/m<sup>3</sup>."
- ❑ **Оценителят** - "При предишни проекти превишението беше около 15%, затова се оттеглих."
- ❑ **Ръководител на отдел "Данни" или инженер по контрола от страна на клиента** - "Нека да разгледаме анализите за последните три търга."

След като анализираме DataFrame от предишни проекти, получаваме:

- Средна действителна покупна цена: 84,80 /m<sup>3</sup>€
- Среден коефициент на превишаване на разходите: +4,7%
- ▲ Препоръчителна норма в оценката: ~ 85 /m<sup>3</sup>€

Такова решение вече няма да се основава на субективни усещания, а на конкретни исторически статистически данни, което помага да се намалят рисковете и да се увеличи валидността на оферата за търга. Анализът на данните от минали проекти се превръща в своеобразен "ограничен тор", от който покълват нови, по-точни решения.



Фигура 7.1-3 Анализът на данни дава отговор на три ключови въпроса: какво се е случило, защо се е случило и какво трябва да се направи по-нататък.

Лицата, вземащи решения, и мениджърите често се сблъскват с необходимостта да работят с разнородни данни и системи, без да разполагат с достатъчно технически познания. В такива ситуации визуализацията, която е една от първите и най-важни стъпки в аналитичния процес, е ключов помощник за разбирането на данните. Тя позволява да се представи информацията във визуална и разбираема форма.

## Визуализиране на данни от: ключ към разбирането и вземането на решения

В днешната строителна индустрия, където данните за проектите се характеризират със сложност и многостепенна структура, визуализацията играе ключова роля. Визуализацията на данни позволява на ръководителите на проекти и инженерите да визуализират сложни модели и тенденции, скрити в големи, разнородни обеми от данни.

Визуализацията на данните от улеснява разбирането на състоянието на даден проект: разпределение на ресурсите, тенденции в разходите или използване на материали. Графиките и диаграмите правят сложната и суха информация достъпна и разбираема, като ви позволяват бързо да идентифицирате ключови области, които се нуждаят от внимание, и да забележите потенциални проблеми.

Визуализацията на данните от не само улеснява интерпретацията на информацията, но е и важна стъпка в аналитичния процес и вземането на информирани управленски решения, като помага да се отговори на въпросите "какво се е случило?" и "как се е случило?" (фигура 2.2-5). (Фигура 2.2-5).

*Графиките са визуални инструменти за решаване на логически задачи [132].*

- Жак Бертин, "Графики и графична обработка на информацията

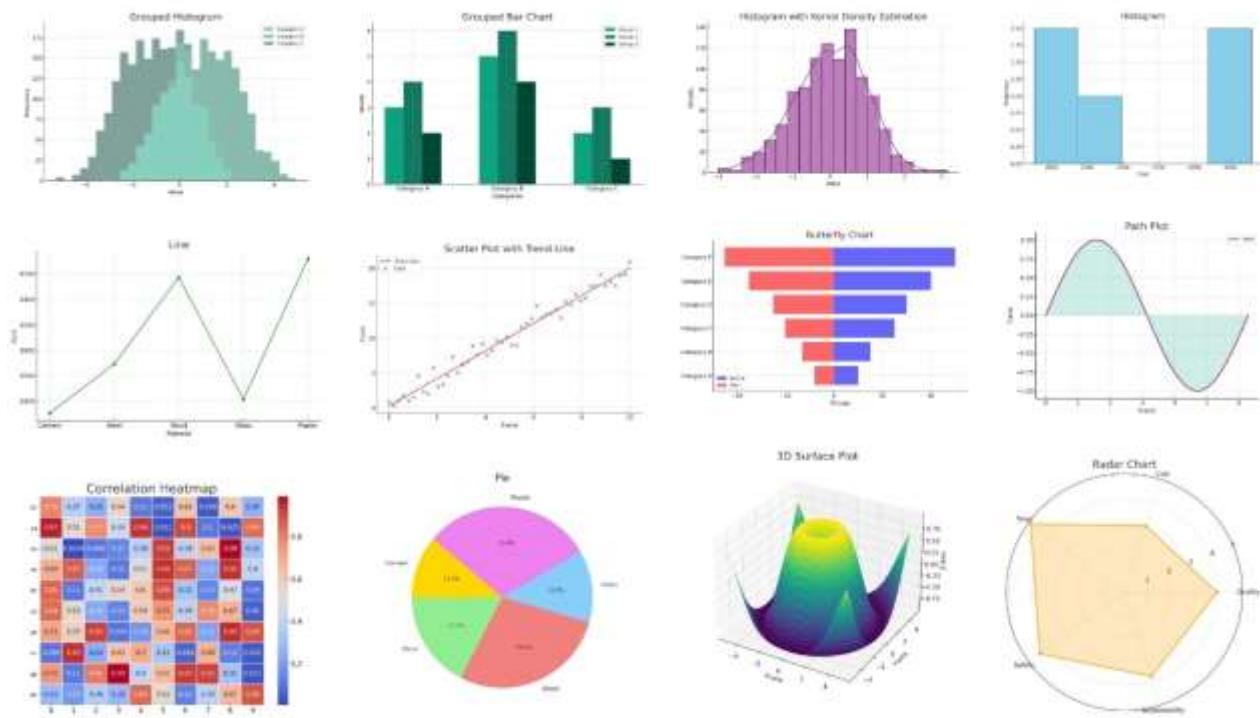
Преди да вземат ключови решения, мениджърите на проекти са по-склонни да използват визуално представяне на данните, отколкото сухи и трудни за тълкуване цифри от електронни таблици или текстови съобщения.

Данните без визуализация са като строителните материали, разхвърляни безразборно на строителната площадка: техният потенциал е неясен. Едва когато се визуализират ясно, като къща от тухли и бетон, тяхната стойност става ясна. Докато къщата не бъде построена, не може да се каже дали купчината материали ще се превърне в малка къщичка, луксозна вила или небостъргач.

Компаниите разполагат с данни от различни системи (от Фигура 1.2-4 до Фигура 2.1-10), финансови транзакции и обширни текстови данни. Използването на тези данни за бизнес ползи обаче често е предизвикателство. В такива ситуации визуализацията се превръща във важен инструмент за предаване на значението на данните, като помага да се представи информацията във формати, които всеки експерт може да разбере, като например информационни таблица, графики и диаграми.

В проучването на PwC "От какво се нуждаят учениците, за да успеят в бързо променящия се свят на бизнеса" (2015 г.) се подчертава [9], че успешните компании отиват отвъд анализа на данни и активно използват интерактивни инструменти за визуализация като графики, инфографики и аналитични табла, за да подпомогнат вземането на решения. Според доклада - визуализацията на данни помага на клиентите да разберат историята, която данните разказват, чрез графики, диаграми, информационни табла и интерактивни модели на данни.

Процесът на преобразуване на информацията във визуални графични форми, като диаграми, графики и схеми, подобрява разбирането и тълкуването на данните от човешкия мозък (фигура 7.1-4). Това позволява на ръководителите на проекти и анализаторите по-бързо да оценяват сложни сценарии и да вземат информирани решения въз основа на визуално разпознаваеми тенденции и модели, а не на интуиция.



**Фиг. 7.1-4** Различните видове визуализации са предназначени да помогнат на човешкия мозък да разбере и осмисли по-добре сухата информация на числата.

Създаването на визуализации от данни и използването на различни безплатни библиотеки за визуализация ще бъдат разгледани по-подробно в следващата глава за ETL -процесите.

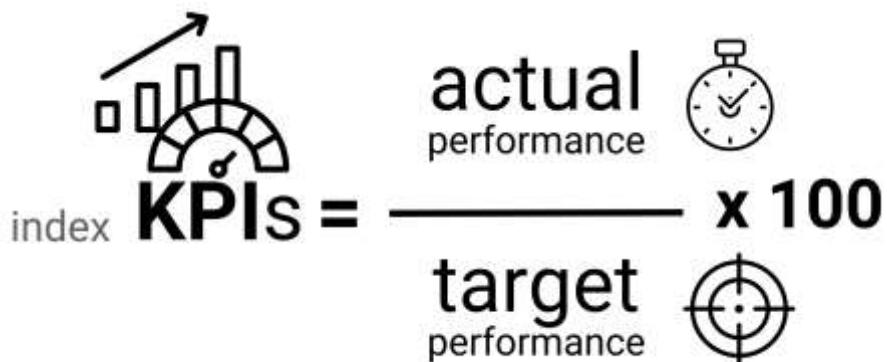
Визуализацията се превръща в неразделен елемент от работата с данни в строителната индустрия - тя помага не само да се "видят" данните, но и да се разбере тяхното значение в контекста на управлениските задачи. Въпреки това, за да бъде визуализацията наистина полезна, е необходимо предварително да се определи какво точно трябва да се визуализира и кои показатели са наистина важни за оценката на изпълнението на проекта. Именно тук влизат в действие показатели за ефективност, като например ключови показатели за ефективност и възвръщаемост на инвестициите. Без тях дори и най-красивите табла за управление рискуват да бъдат просто "информационен шум".

## Ключови показатели за ефективност и възвръщаемост на инвестициите

В съвременната строителна индустрия управлението на показателите за ефективност (KPI и ROI) и тяхното визуализиране чрез отчети и информационни табла играят ключова роля за подобряване на производителността и ефективността на управлението на проекти.

Както във всеки бизнес, така и в строителството е необходимо ясно да се определят показателите, по които се измерват успехът, възвръщаемостта на инвестициите и ефективността. При получаването на данни за различни процеси една организация, която се ръководи от данните, трябва първо да се научи да идентифицира **KPI** (**Key Performance Indicators**) - количествени показатели, които отразяват степента на постигане на стратегическите и оперативните цели.

За да се изчисли KPI, обикновено се използва формула (фигура 7.1-5), която включва действителните и планираните показатели. Например, за да изчислите индивидуален KPI за проект, служител или процес, разделете действителните показатели на планираните показатели и умножете резултата по 100%.



Фигура 7.1-5 Ключовите показатели за ефективност (KPI) се използват за измерване на успеха на даден проект или процес при постигането на ключови цели.

На ниво сайт могат да се използват по-подробни ключови показатели за ефективност:

- **График на ключовите етапи** (фундиране, монтаж, довършителни работи) - позволява да се контролира спазването на работните планове.
- **Процент на преразход на материали** - помага за управление на снабдяването и свеждане до минимум на загубите.
- **Брой на непланираните престои на машините** - отразява се на производителността и разходите.

Изборът на погрешни показатели може да доведе до погрешни решения "какво да се прави?" (Фигура 2.2-5). Например, ако компанията се фокусира само върху разходите на квадратен метър, но не вземе предвид разходите за преустройства, икономиите от материали могат да доведат до по-лошо качество и по-високи разходи в бъдещи проекти.

Когато поставяте цели, е важно да сте наясно какво се измерва. Неясната формулировка води до неправилни заключения и усложнява контрола. Нека разгледаме примери за успешни и неуспешни KPI в строителството.

#### **Добри ключови показатели за ефективност:**

- ☒ "До края на годината намалете процента на пребоядисване с 10%."
- ☒ "Увеличаване на скоростта на монтаж на фасади с 15% без компромис с качеството до следващото тримесечие"
- ☒ "Намаляване на времето за престой на машините с 20% чрез оптимизиране на работните графици до края на годината"

Тези показатели са ясно измерими, имат конкретни стойности и срокове.

#### **Лоши ключови показатели за ефективност:**

- ☒ "Ще строим по-бързо" (Колко по-бързо? Какво означава "по-бързо"?).
- ☒ "Ще подобрим качеството на бетонната работа" (Как точно се измерва качеството?)
- ☒ "Ще подобрим взаимодействието с изпълнителите на място" (Какви критерии ще покажат подобрението?)

Добрият ключов показател за ефективност е този, който може да бъде измерен и обективно оценен. В строителството това е особено важно, тъй като без ясни показатели е невъзможно да се следи изпълнението и да се постигнат стабилни резултати.

В допълнение към KPI съществува допълнителен показател за оценка на ефективността на инвестициите: **ROI (Return on Investment)** - показател за възвръщаемост на инвестициите, отразяващ съотношението между печалбата и инвестираните средства. Възвръщаемостта на инвестициите помага да се прецени дали въвеждането на нови методи, технологии или инструменти е оправдано: от цифровите решения и автоматизацията (напр. фигура 7.3-2) до използването на нови строителни материали. Този показател помага да се вземат информирани решения за по-нататъшни инвестиции въз основа на реалното им въздействие върху рентабилността на бизнеса

В контекста на управлението на строителни проекти ROI (възвръщаемост на инвестициите) може да се използва като един от ключовите показатели за ефективност (KPI), ако целта на компанията е да измери възвръщаемостта на инвестициите в даден проект, технология или подобреие на процеса. Например, ако се внедрява нова техника за управление на строителството, ROI може да покаже доколко тя е подобрila рентабилността.

Редовното измерване на ключовите показатели за ефективност и възвръщаемост на инвестициите въз основа на данни, събрани от различни източници, като например потреблението на материали, часовете труд и разходите, дава възможност на ръководството на проекта ефективно да управлява ресурсите и да взема бързи решения. Съхраняването на тези данни в дългосрочен план позволява да се анализират бъдещите тенденции и да се оптимизират процесите.

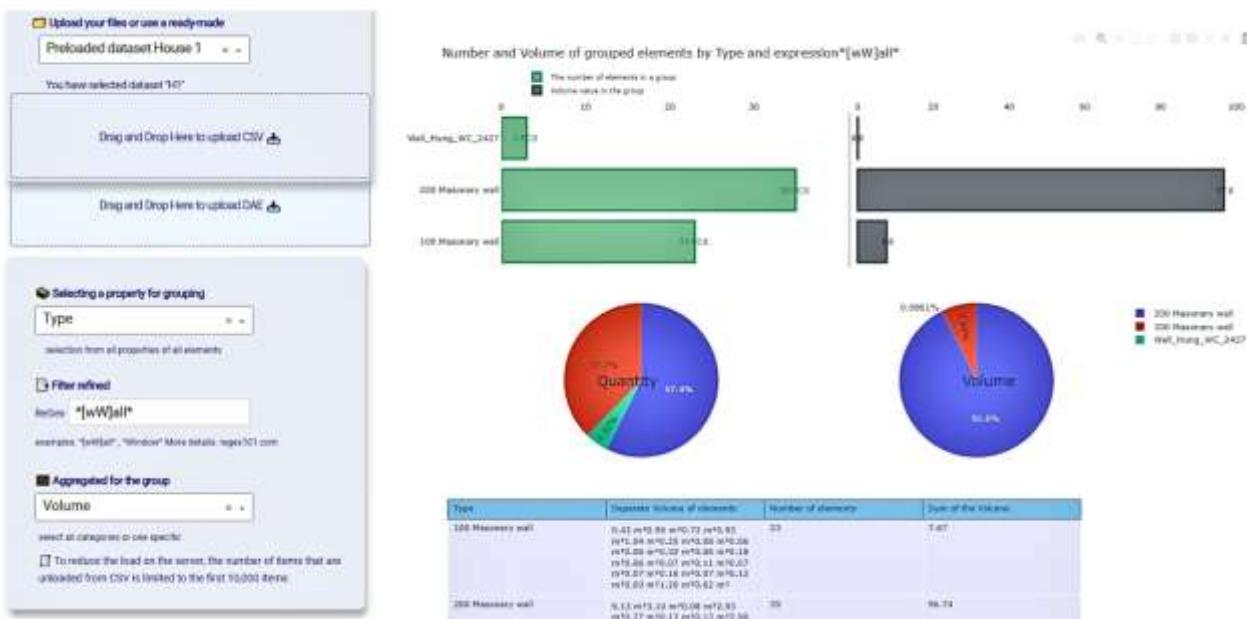
За визуализиране на ключови показатели за ефективност, възвръщаемост на инвестициите и

други показатели се използват различни диаграми и графики, които обикновено се обединяват в информационни табла.

## Информационни табла и табла за управление: визуализиране на показатели за ефективно управление

За визуализиране на показателите и метриките се използват разнообразни диаграми и графики, които обикновено се обединяват във витрини с данни и информационни табла. Тези табла представят централизиран поглед върху състоянието на проекта или части от него, като показват ключови показатели (в идеалния случай в реално време). Актуалните и непрекъснато обновявани табла позволяват на екипа да реагира бързо на промените.

**Информационни табла** са инструменти, които визуализират количествените оценки, като ги правят лесно достъпни и разбираеми за всички участници в проекта.



Фиг. 7.1-6 Управлението на ключовите показатели за ефективност (KPI) и визуализирането им чрез информационни табла е от ключово значение за подобряване на производителността и ефективността на проекта.

Ето няколко примера за популярни инструменти, с които можете да създавате информационни табла:

- **Power BI** е инструмент на Microsoft за създаване на интерактивни отчети и информационни табла.
- **Tableau и Google Data Studio** са мощни инструменти за визуализиране на данни и създаване на информационни табла, без да се налага да пишете код.

- **Plotly** (Фигура 7.1-6, Фигура 7.2-12) е библиотека за създаване на интерактивни графики, а Dash е рамка за създаване на уеб приложения за анализ на данни. Те могат да се използват в комбинация за създаване на интерактивни информационни табла.
- **Многобройни библиотеки за Python** (Фигура 7.2-9 - Фигура 7.2-11) - Python разполага с много безплатни библиотеки с отворен код за визуализация на данни, като Matplotlib, Seaborn, Plotly, Bokeh и други. Те могат да се използват за създаване на графики и интегрирането им в уеб приложение с помощта на рамки като Flask или Django.
- **Библиотеки на JavaScript**: позволява ви да създавате интерактивни табла с помощта на библиотеки на JavaScript с отворен код, като D3.js или Chart.js, и да ги интегрирате в уеб приложение.

За да можете да оценявате ключовите показатели за ефективност и да създавате информационни табла, се нуждаете от актуални данни и ясен график за събиране и анализиране на информацията.

Като цяло ключовите показатели за ефективност, възвръщаемостта на инвестициите и информационните табла в строителния бранш формират основата за аналитичен подход към управлението на проекти. Те не само помагат да се наблюдава и оценява текущото състояние, но и предоставят ценни сведения за бъдещите процеси на планиране и оптимизация - процеси, които зависят пряко от интерпретирането на данните и задаването на правилните и навременни въпроси.

## Анализ на данни и изкуството да се задават въпроси

Интерпретацията на данните е последният етап от анализа, в който информацията придобива смисъл и започва да "говори". Тук се формулират отговорите на ключовите въпроси: "Какво да се направи?" и "Как да се направи?" (Фигура 2.2-5). Този етап позволява да се обобщят резултатите, да се идентифицират закономерности, да се установят причинно-следствени връзки и да се направят заключения въз основа на визуализация и статистически анализ.

*Може би не е далеч времето, когато ще се разбере, че за да стане човек пълноценен гражданин на една от големите нови сложни световни държави, които сега се развиват, е необходимо да умееш да изчислява, да мисли в термини на средни стойности, максимуми и минимуми, както сега е необходимо да може да чете и пише [133].*

- Самуел С. Уилкс, цитиран в президентската реч пред Американската статистическа асоциация през 1951 г.

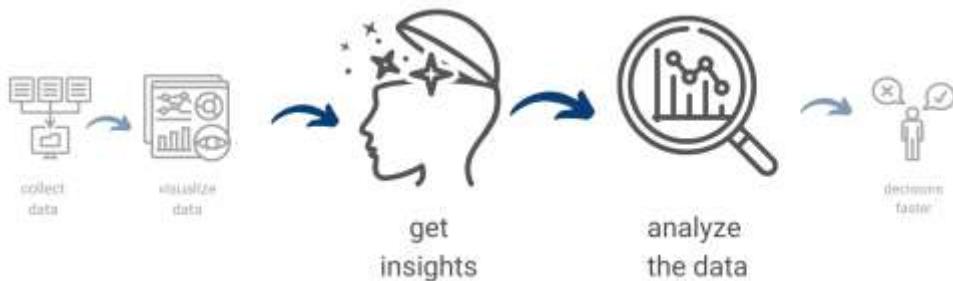
Според доклада "Анализ на данни и изкуствен интелект при изпълнението на правителствени проекти" (2024 г.), публикуван от правителството на Обединеното кралство [83], прилагането на анализ на данни и изкуствен интелект (AI) може значително да подобри процесите на управление на проекти, като увеличи точността на прогнозиране на времето и разходите, както и да намали риска и несигурността. В документа се подчертава, че публичните организации, които използват усъвършенствани аналитични инструменти, постигат по-високи резултати в инфраструктурните инициативи.

Съвременният строителен бизнес, работещ в силно конкурентната среда на четвъртата индустриална революция с нисък марж, може да се сравни с военни операции. Тук оцеляването и успехът на фирмата зависят от бързината на получаване на ресурси и качествена информация - а оттам и от навременното и информирано вземане на решения (фиг. 7.1-7).

Ако визуализацията на данни е "разузнавателната информация", която дава общ поглед, то анализът на данни е "амуницията", необходима за действие. Той дава отговор на въпросите: какво да се направи и как да се направи, като по този начин се формира основата за придобиване на конкурентно предимство на пазара.

Анализът превръща разнородните данни в структурирана и смислена информация, на която да се базират решенията.

Задачата на анализаторите и мениджърите е не само да интерпретират информацията, но и да предлагат информирани решения, да идентифицират тенденции, да определят връзките между различните видове данни и да ги категоризират в съответствие с целите и спецификата на проекта. Използвайки инструменти за визуализация и методи за статистически анализ, те превръщат данните в стратегически актив за компанията.



**Фиг. 7.1-7 Анализът на данните е този, който в крайна сметка превръща събраната информация в източник за вземане на решения.**

За да се вземат наистина информирани решения в процеса на анализ, е необходимо да се научите как правилно да формулирате въпросите, които се задават на данните. Качеството на тези въпроси пряко влияе върху дълбочината на получените прозрения и, като следствие, върху качеството на управленските решения.

*Миналото съществува само дотолкова, доколкото присъства в днешните записи. А какво представляват тези записи, се определя от въпросите, които задаваме. Няма друга история освен тази [134].*

- Джон Арчибалд Уилър, физик 1982 г.

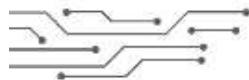
Изкуството да се задават задълбочени въпроси и да се мисли критично е основно умение при работа с данни. Повечето хора са склонни да задават прости, повърхностни въпроси, чийто отговор не изисква особени усилия. Истинският анализ обаче започва със сmisлени и обmisлени въпроси, които могат да разкрият скрити връзки и причинно-следствени връзки в информацията, която може да е скрита зад множество пластове разсъждения.

Според проучването "Data-Driven Transformation: Accelerating at Scale Now" (BCG, 2017 г.) [135] успешната цифрова трансформация изисква инвестиции в аналитични способности, програми за управление на промяната и съгласуване на бизнес целите с ИТ инициативите. Компаниите, които създават култура, основана на данни, трябва да инвестират в способности за анализ на данни и да стартират програми за управление на промените, за да внушат ново мислене, поведение и начини на работа.

Без инвестиции в развитието на аналитична култура, подобряване на инструментите за обработка на данни и обучение на специалисти, компаниите ще продължат да рискуват да вземат решения въз основа на остатяла или непълна информация - или да разчитат на субективното мнение на мениджърите на HiPPO (фигура 2.1-9).

Осъзнаването на значимостта и необходимостта от постоянно актуализиране на анализите и информационните табла неизбежно кара ръководството да разбере важността на автоматизирането на аналитичните процеси. Автоматизацията повишава скоростта на вземане на решения, намалява влиянието на човешкия фактор и осигурява релевантност на данните. С експоненциалното нарастване на обема на информацията скоростта се превръща не просто в конкурентно предимство, а в ключов фактор за устойчив успех.

Автоматизацията на процесите на анализ и обработка на данни като цяло е неразрывно свързана с темата за ETL (Extract, Transform, Load). Точно както в процеса на автоматизация трябва да трансформираме данните, в процеса ETL данните се извличат от различни източници, трансформират се според необходимите изисквания и се зареждат в целевите системи за по-нататъшно използване.



## ГЛАВА 7.2.

### ПОТОК НА ДАННИ БЕЗ РЪЧНИ УСИЛИЯ: ЗАЩО Е НЕОБХОДИМ ETL

#### Автоматизация на ETL: намаляване на разходите и ускоряване на обработката на данни

Когато ключовите показатели за ефективност (KPI) престанат да нарастват въпреки увеличаването на обема на данните и размера на екипа, ръководството на компанията неизбежно осъзнава необходимостта от автоматизиране на процесите. Рано или късно това осъзнаване се превръща в стимул за стартиране на комплексна автоматизация, чиято основна цел е да се намали сложността на процесите, да се ускори обработката и да се намали зависимостта от човешкия фактор.

Според проучването на McKinsey "Как да изградим архитектура за данни, за да стимулираме иновациите - днес и утре" (2022 г.) [136], компаниите, които използват архитектури за поточни данни, получават значително предимство, тъй като могат да анализират информацията в реално време. Поточните технологии позволяват директен анализ на съобщения в реално време и прилагане на прогнозна поддръжка в производството чрез анализ на данни от сензори в реално време.

Опростяването на процесите е автоматизация, при която традиционните ръчни функции се заменят с алгоритми и системи.

Въпросът за автоматизацията или по-скоро за "свеждането до минимум на ролята на человека в обработката на данни" е необратим и много чувствителен процес за всяка компания. Специалистите в която и да е професионална област често се колебаят дали да разкрият напълно своите методи и тънкости на работа пред колегите си оптимизатори, осъзнавайки риска да загубят работата си в бързо развиващата се технологична среда.

Ако искате да си спечелите врагове, опитайте се да промените нещата [137].

- Удру Уилсън, реч пред конгрес на търговците, Детройт, 1916 г.

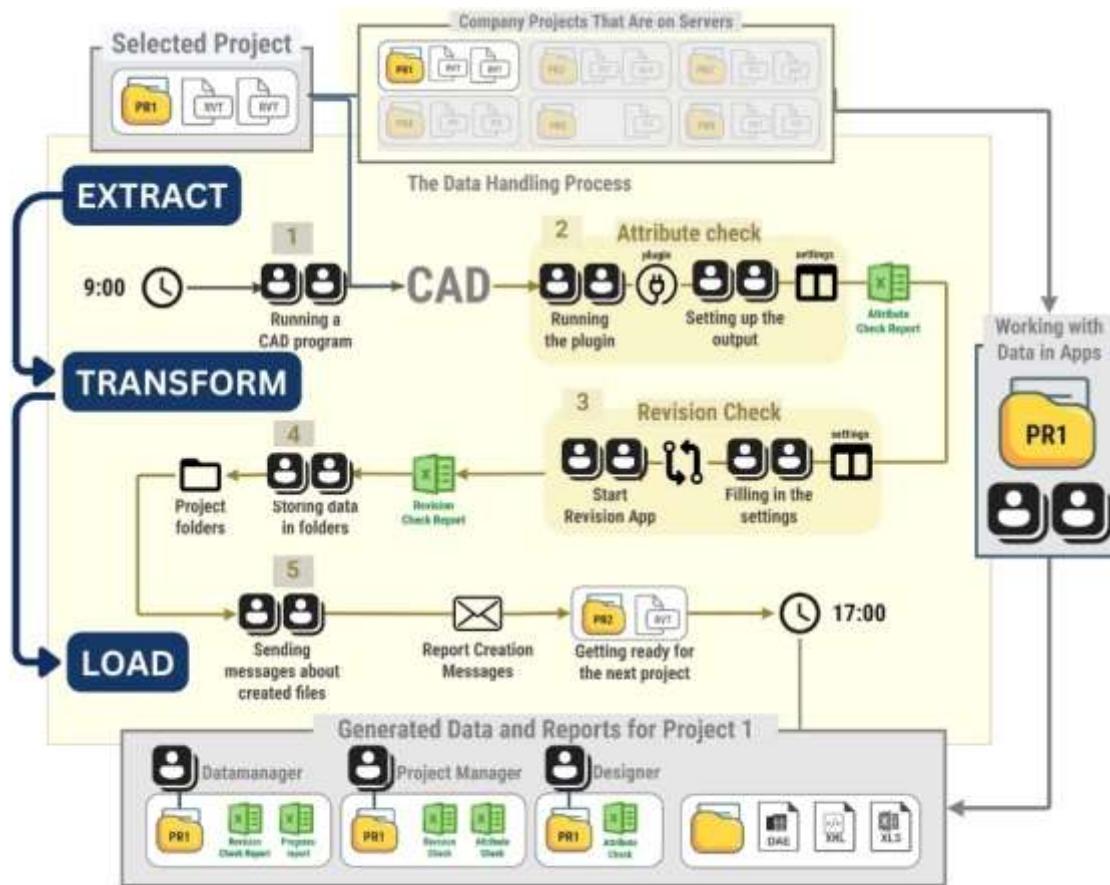
Въпреки очевидните ползи от автоматизацията, много компании все още използват голям дял ръчен труд в ежедневните си практики, особено в областта на инженерните данни. За да илюстрираме настоящата ситуация, нека разгледаме типичен пример за последователна обработка на данни в рамките на такива процеси.

Ръчната обработка на данни може да се илюстрира с примера на взаимодействие с информация, получена от CAD бази данни. Традиционната обработка на данни ("ръчен" ETL-процес) в CAD (BIM) отделите за създаване на атрибутни таблици или за създаване на документация въз основа на проектни данни се извършва в следния ред (фиг. 7.2-1):

1. Ръчно **извлечане (Extract)**: потребителят отваря проекта ръчно - чрез стартиране на CAD

приложението (BIM) (фиг. 7.2-1 стъпка 1).

2. **Проверка:** следващата стъпка обикновено включва ръчно стартиране на няколко приставки или помощни приложения на, за да се подготвят данните и да се оцени тяхното качество (фиг. 7.2-1 стъпка 2-3).
3. Ръчно **преобразуване (Transform):** след подготовката започва обработката на данните, което изисква ръчна работа с различни софтуерни инструменти, в които данните се подготвят за качване (фиг. 7.2-1 стъпка 4).
4. Ръчно **качване (Load):** ръчно качване на конвертирани данни във външни системи, формати на данни и документи (фиг. 7.2-1 стъпка 5).



Фиг. 7.2-1 Традиционната ръчна обработка на ETL е ограничена от желанията и физическите възможности на отделния техник.

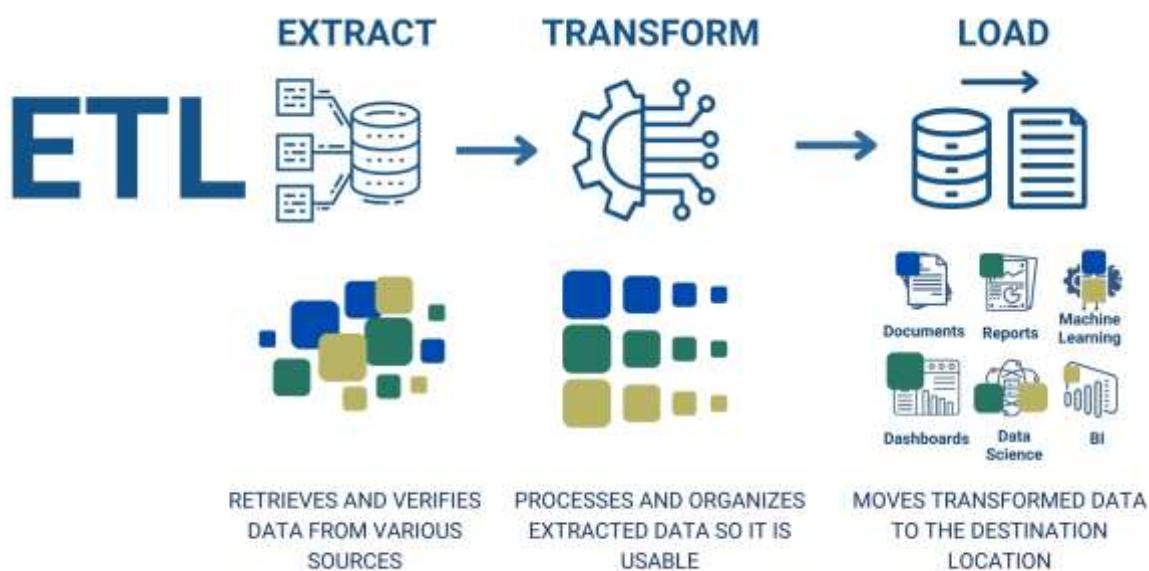
Такъв работен процес е пример за класически ETL -процес - извлечане, преобразуване и зареждане (ETL). За разлика от други индустрии, където автоматичните ETL конвейери отдавна са стандарт, в строителната индустрия все още преобладава ръчният труд, който забавя процесите и увеличава разходите

**ETL (Extract, Transform, Load)** е процесът на извличане на данни от различни източници, трансформирането им в желания формат и зареждането им в целевата система за по-нататъшен анализ и използване.

ETL е процес, който обозначава три ключови компонента на обработката на данни: Извличане, Трансформиране и Зареждане (фигура 7.2-2):

- **Извличане** - извличане на данни от различни източници (файлове, бази данни, API).
- **Трансформиране** - почистване на данни, агрегиране, нормализиране и логическа обработка.
- **Зареждане** - зареждане на структурирана информация в склад за данни, отчет или BI система.

По-рано в книгата концепцията за ETL беше засегната само от време на време: при преобразуването на неструктурен сканиран документ в структуриран табличен формат (фиг. 4.1-1), в контекста на формализирането на изискванията с цел систематизиране на възприемането на жизнените и бизнес процесите (фиг. 4.4-20) и при автоматизирането на валидирането и обработката на данни от CAD решения. Нека сега да разгледаме по-подробно ETL в контекста на типичните работни процеси.



Фигура 7.2-2 ETL автоматизира повтарящи се задачи за обработка на данни.

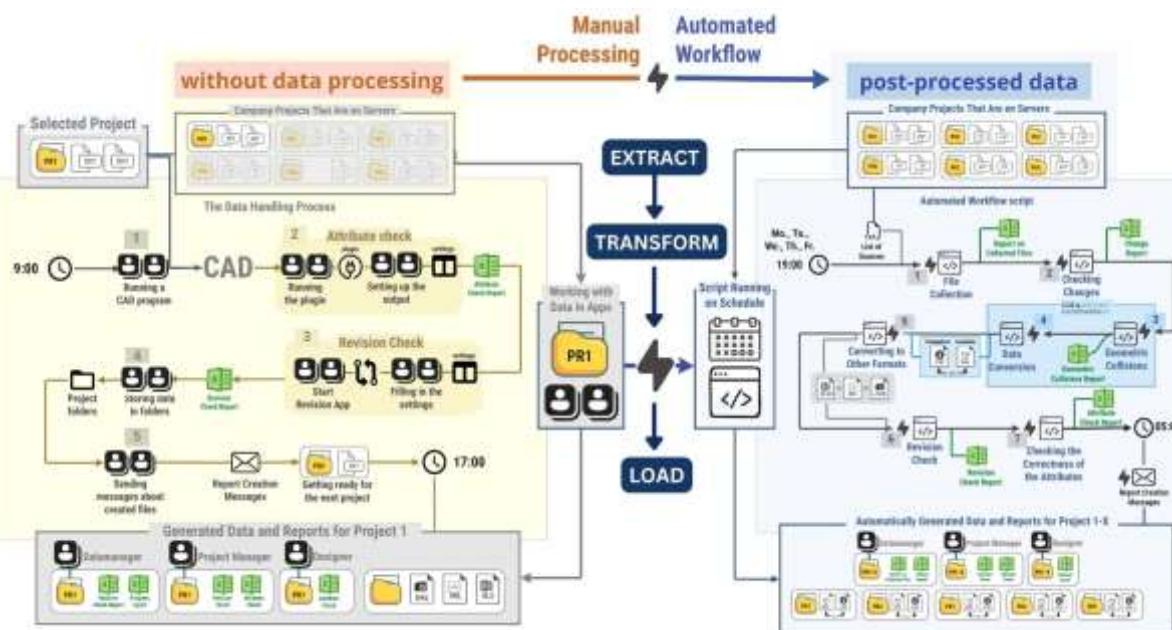
Ръчен или полуавтоматичен ETL -процесът предполага мениджър или техник, който управлява всички стъпки ръчно - от събирането на данни до генерирането на отчети. Такъв процес отнема значително време, особено когато работното време е ограничено (напр. от 9:00 до 17:00 ч.).

Компаниите често се стремят да решат проблема с ниската ефективност и бавната скорост, като закупуват модулни интегрирани решения (ERP, PMIS, CPM, CAFM и т.н.), които след това се доразвиват от външни доставчици и консултанти. Но тези доставчици и външни разработчици често се

превръщат в критична точка на зависимост: техните технически ограничения пряко влияят на ефективността на цялата система и на бизнеса като цяло, както е описано подробно в предишните глави за патентованите системи и формати. Проблемите, породени от фрагментацията и зависимостта, бяха разгледани подробно в главата "Как строителните предприятия потъват в хаоса от данни".

Ако компанията не е готова да внедри голяма модулна платформа от някой от доставчиците, тя започва да търси алтернативни начини за автоматизация. Един от тях е да се разработят собствени модулни отворени ETL-конвейери, в които всеки етап (извлечане, преобразуване, валидиране, зареждане) се реализира като скриптове, изпълнявани по график.

В автоматизираната версия на същия работен процес ETL (фиг. 7.2-1) работният процес изглежда като модулен код, който започва с обработката на данни и превеждането им в отворена структурирана форма. След като се получат структурираните данни, автоматично, по график, се изпълняват различни скриптове или модули за проверка на промените, трансформиране и изпращане на съобщения (фиг. 7.2-3).



Фиг. 7.2-3 Вляво е ръчната обработка, а вдясно - автоматичен процес, който за разлика от традиционната ръчна обработка не е ограничен от възможностите на потребителя.

В автоматизирания работен процес обработката на данните се опростява чрез предварителна обработка на данните ETL: структуриране и обединяване.

При традиционните методи за обработка специалистите работят с данните "такива, каквите са" - както са извлечени от системите или софтуера. При автоматизираните процеси, напротив, данните често първо преминават през ETL -payplane, където се привеждат в последователна структура и формат, подходящи за по-нататъшно използване и анализ.

Нека вземем практически пример за ETL, демонстриращ процеса на валидиране на таблицата с данни, описан в главата "Валидиране на данни и резултати от валидирането" (фигура 4.4-13). За целта използваме библиотеката Pandas в комбинация с LLM за автоматизиран анализ на данни и процеси на обработка.

## ETL Извлечение: събиране на данни

Първият етап от процеса ETL - Extract) - започва с писане на код за събиране на набори от данни, които да бъдат допълнително проверени и обработени. За тази цел ще сканираме всички папки на производствения сървър, ще съберем документи с определен формат и съдържание, след което ще ги преобразуваме в структуриран вид. Този процес е разгледан подробно в главите "Конвертиране на неструктурирани и текстови данни в структурирана форма" и "Конвертиране на CAD данни (BIM) в структурирана форма" (фиг. 4.1-1 - фиг. 4.1-12).



Фиг. 7.2-4 Конвертиране на CAD данни (BIM) в една голяма рамка за данни, която ще съдържа всички секции на проекта.

Като илюстративен пример използваме стъпката за зареждане на данни Extract и получаваме таблица на всички проекти на CAD- (BIM-) (фиг. 7.2-4) използва конвертори с възможност за обратен инженеринг [138] за форматите RVT и IFC, за да получим структурирани таблици от всички проекти и да ги обединим в една голяма таблица DataFrame.

```

1 import os
2 import subprocess
3 import time
4 import pandas as pd
5
6 path_conv = r'C:\DDC_2023\' # Where RvtExporter.exe>IfcExporter.exe is located
7 path = r'C:\IFCprojects\' # Where Revit|IFC files are stored
8
9 def convert_and_wait(path_conv, exporter_name, file_path, extension):
10    # Start the conversion process
11    subprocess.Popen([os.path.join(path_conv, exporter_name),
12                      file_path], cwd=path_conv)
13    output_file = os.path.join(path,
14                               f"{os.path.splitext(os.path.basename(file_path))[0]}_{extension}.xlsx")
15
16    # Conversion process for RVT and IFC files
17    for file in os.listdir(path):
18        full_path = os.path.join(path, file)
19        if file.endswith('.ifc'):
20            convert_and_wait(path_conv, 'IfcExporter.exe', full_path, 'ifc')
21        elif file.endswith('.rvt'):
22            convert_and_wait(path_conv, 'RvtExporter.exe', full_path, 'rvt')
23
24    # Combine converted Excel files into one dataframe
25 df = pd.concat([pd.read_excel(os.path.join(path, f)) for f in os.listdir(path)
26 if f.endswith('.xlsx')], ignore_index=True)

```

Фиг. 7.2-5 Конвертиране с помощта на Python код и SDK инструмент за обратно инженерство за RVT и IFC файлове в една голяма структурирана (df) DataFrame.

Pandas DataFrame може да зарежда данни от различни източници, включително CSV текстови файлове, електронни таблици на Excel, JSON - и XML - файлове, големи формати за съхранение на данни, като Parquet и HDF5, и от MySQL, PostgreSQL, SQLite, Microsoft SQL Server, Oracle и други бази данни. Освен това Pandas поддържа зареждане на данни от API, уебстраници, облачни услуги и системи за съхранение като Google BigQuery, Amazon Redshift и Snowflake.

- ❷ За да напишете код за свързване и събиране на информация от бази данни, изпратете подобна текстова заявка в чат стаята на LLM (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN или някоя друга):

Моля, напишете пример за свързване към MySQL и конвертиране на данни в ↴

■ Отговор на LLM :

```

1 import pandas as pd
2 import mysql.connector
3
4 # Using a context manager to ensure
5 # that the connection is closed automatically
6 with mysql.connector.connect(host='hostname', database='dbname',
7 user='username', password='password') as conn:
8     df = pd.read_sql('SELECT * FROM tablename LIMIT 10', conn)

```

Run in IDE

**Фигура 7.2-6 Пример за свързване чрез Python към база данни MySQL и импортиране на данни от базата данни MySQL в DataFrame .**

Полученият код (фиг. 7.2-5, фиг. 7.2-6) може да бъде стартиран в една от популярните IDE (интегрирани среди за разработка), които споменахме по-горе, в офлайн режим: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с PyDev плъгин, Thonny, Wing IDE, IntelliJ IDEA с Python плъгин, JupyterLab или популярни онлайн инструменти: Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Като заредихме многоформатните данни в променливата "df" (Фигура 7.2-5 - ред 25; Фигура 7.2-6 - ред 8), преобразувахме данните във формата Pandas DataFrame, една от най-популярните структури за обработка на данни, която представлява двумерна таблица с редове и колони. За други формати за съхранение, използвани в ETL - процесите, като Parquet, Apache ORC, JSON, Feather, HDF5, както и за съвременните складове за данни, ще говорим повече в главата "Съхранение и управление на данни в строителната индустрия" (фигура 8.1-2).

След етапа на извлечане и структуриране на данните (Extract) се формира единен масив от информация (фиг. 7.2-5, фиг. 7.2-6), готов за по-нататъшна обработка. Преди обаче да се заредят тези данни в целевите системи или да се използват за анализ, е необходимо да се гарантира тяхното качество, цялостност и съответствие с определените изисквания. На този етап се извършва преобразуване на данните (Transform) - ключова стъпка, която гарантира надеждността на последващите заключения и решения.

## ETL Трансформиране: прилагане на правила за валидиране и трансформиране

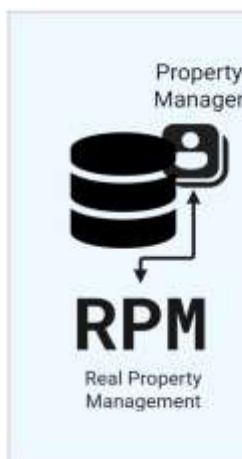
На етапа Transform данните се обработват и трансформират. Този процес може да включва проверка на коректността, нормализация, попълване на липсващи стойности и валидиране с помощта на автоматизирани инструменти.

Според проучването на PwC "Data-Driven. От какво се нуждаят студентите, за да успеят в бързо променящия се бизнес свят" (2015 г.) [9], съвременните одиторски фирми се отказват от случайната проверка на данни и преминават към анализ на огромни количества информация с помощта на автоматизирани инструменти. Този подход дава възможност не само да се идентифицират несъответствия в отчетността, но и да се предлагат препоръки за оптимизиране на бизнес процесите.

В строителството подобни методи могат да се използват например за автоматично валидиране на проектни данни, контрол на качеството на строителството и оценка на работата на изпълнителите. Един от инструментите за автоматизиране и ускоряване на обработката на данни е използването на регулярни изрази (RegEx) във фазата на преобразуване на данните (Transform) на процеса ETL. RegEx позволява ефективно да се валидират низове от данни, да се откриват несъответствия и да се гарантира целостта на информацията с минимално потребление на ресурси. Повече подробности за RegEx (фиг. 4.4-7) бяха разгледани в главата "Превеждане на изискванията в структуриран вид".

Нека разгледаме практически пример: в системата за управление на обекти на собственост (RPM) управителят задава изискванията за ключовите атрибути на обектите (фиг. 7.2-7). На етапа на преобразуване е необходимо да се валидират следните параметри:

- проверка на форматите на идентifikатора на обекта (атрибут "ID")
- контрол на стойностите на гаранционния срок за замяна (атрибут "Гаранционен срок")
- проверка на цикъла на подмяна на елементите (атрибут "Изисквания за поддръжка").



**Property Manager:  
Long-term Management**

ID	Element	Warranty Period	Replacement Cycle	Maintenance Requirements
W-NEW	Window	-	20 years	Annual Inspection
W-DLD1	Window	8 years	15 years	Biannual Inspection
W-DLD2	Window	8 years	15 years	Biannual Inspection
D-122	Door	15 years	25 years	Biennial Varnishing

Фигура 7.2-7 Проверка на качеството започва със задаване на изисквания за атрибутите и техните гранични стойности.

За да зададете гранични стойности за валидиране на параметъра, например, да предположим, че от нашия опит знаем, че валидните стойности за атрибута "ID" могат да включват само символни низове "W-NEW", "W-OLD1" или "D-122" или подобни стойности, при които първият символ е буква, последвана от тире, и след това три буквени символа "NEW", "OLD" или всяко трицифreno число (фигура 7.2-7). Следният регулярен израз (RegEx) може да се използва за валидиране на тези идентификатори:

```
^W-NEW$ | ^W-OLD[0-9]+$ | ^D-1[0-9]{2}$
```

Този шаблон гарантира, че всички идентификатори в данните отговарят на зададените критерии. Ако някоя стойност не премине успешно проверката, системата регистрира грешка. За да създадете Python -код за преобразуване на данните и да използвате получените данни за създаване на таблица с резултати, просто формулирайте заявка в чата LLM.

❷ Заявка за текст до LLM:

Напишете код за валидиране на колоните на DataFrame с помощта на регулярни изрази, който проверява идентификатори във формат "W-NEW" или "W-OLD" чрез RegEx, енергийна ефективност с букви от "A" до "G", гаранционен период и цикъл на подмяна с числови стойности в години ↴

💡 Отговор на LLM :

```

1 # Regex
2 import re
3
4 # Define regex patterns for each column
5 patterns = {
6     'Window_ID': r'^[W-NEW|M-OLD]{0-9}+[D-1]{0-9}[2]{0-9}$',
7     'Energy_Efficiency': r'^[A-G]{1-4}+$',
8     'Warranty_Period': r'^[d]{1-2}[years]$', 
9     'Replacement_Cycle': r'^[d]{1-2}[years]$'
10 }
11
12 # Define a verification function using regex
13 def verify_data(column, pattern):
14     # Apply the regex pattern to the series and return the boolean results
15     return column.apply(lambda x: bool(re.match(pattern, str(x))))
16
17 # Apply the verification function to the DataFrame for each column with its respective pattern
18 for column, pattern in patterns.items():
19     df['verified_' + column] = verify_data(df[column], pattern)
20

```

**Фигура 7.2-8 Кодът автоматизира процеса на валидиране чрез прилагане на RegEx шаблони към колоните с параметри на рамката с данни .**

Горният, автоматично генериран Python -код (Фигура 7.2-8), използва библиотеката "re" (RegEx регулярни изрази), за да дефинира функция, която проверява всеки атрибут на елемент от данни в DataFrame. За всяка посочена колона (атрибут) функцията прилага RegEx шаблон, за да провери дали всеки запис съответства на очаквания формат, и добавя резултатите като нови стойности (False/True) към новата колона с атрибути на DataFrame.

Такова автоматизирано валидиране осигурява формално съответствие на данните и може да се използва като част от системата за контрол на качеството по време на фазата на преобразуване.

След успешното завършване на етапа на трансформиране и проверката на качеството данните са готови за качване в целевите системи. Трансформирани и валидирани данни могат да бъдат качени в CSV, JSON, Excel, бази данни и други формати за по-нататъшно използване. В зависимост от задачата резултатите могат да бъдат представени и в доклади, графики или аналитични табла.

## ETL Load: Визуализирайте резултатите от в диаграми и графики

След приключване на етапа Transform, когато данните са структурирани и проверени, последният етап е Load, при който данните могат да бъдат заредени в целевата система и визуализирани за

анализ. Визуалното представяне на данните позволява бързо да се установят отклоненията, да се анализират разпределенията и да се съобщят основните заключения на всички участници в проекта, включително и на тези без техническа подготовка.

Вместо да представяме информацията под формата на таблици и числа, можем да използваме инфографики, графики и информационни табла (dashboards). Един от най-разпространените и гъвкави инструменти за визуализиране на структурирани данни в Python е библиотеката Matplotlib (Фигура 7.2-9, Фигура 7.2-10). Тя ви позволява да създавате статични, анимирани и интерактивни диаграми и поддържа широк набор от типове диаграми.

- За да визуализирате резултатите от проверката на атрибутите от системата RPM (фигура 7.2-7), можете да използвате следната заявка към езиковия модел:

Напишете код за визуализиране на данните от DataFrame, по-горе (фигура 7.2-7), с хистограма за резултатите, която да показва честотата на грешките в атрибута `✉`

- 💡 Отговорът на LLM под формата на код и готова визуализация директно в чата на LLM на резултатите от изпълнението на кода:

```

1 # Re-importing necessary libraries for visualization
2 import matplotlib.pyplot as plt
3 import seaborn as sns
4
5 # Visualization 1: Bar Chart
6 plt.figure(figsize=(10, 6))
7 df_visual.plot(kind='bar', stacked=True, color=['green', 'red'])
8 plt.title('Data Verification Summary - Bar Chart')
9 plt.xlabel('Data Categories')
10 plt.ylabel('Count')
11 plt.xticks(rotation=45)
12 plt.tight_layout()
13 plt.show()

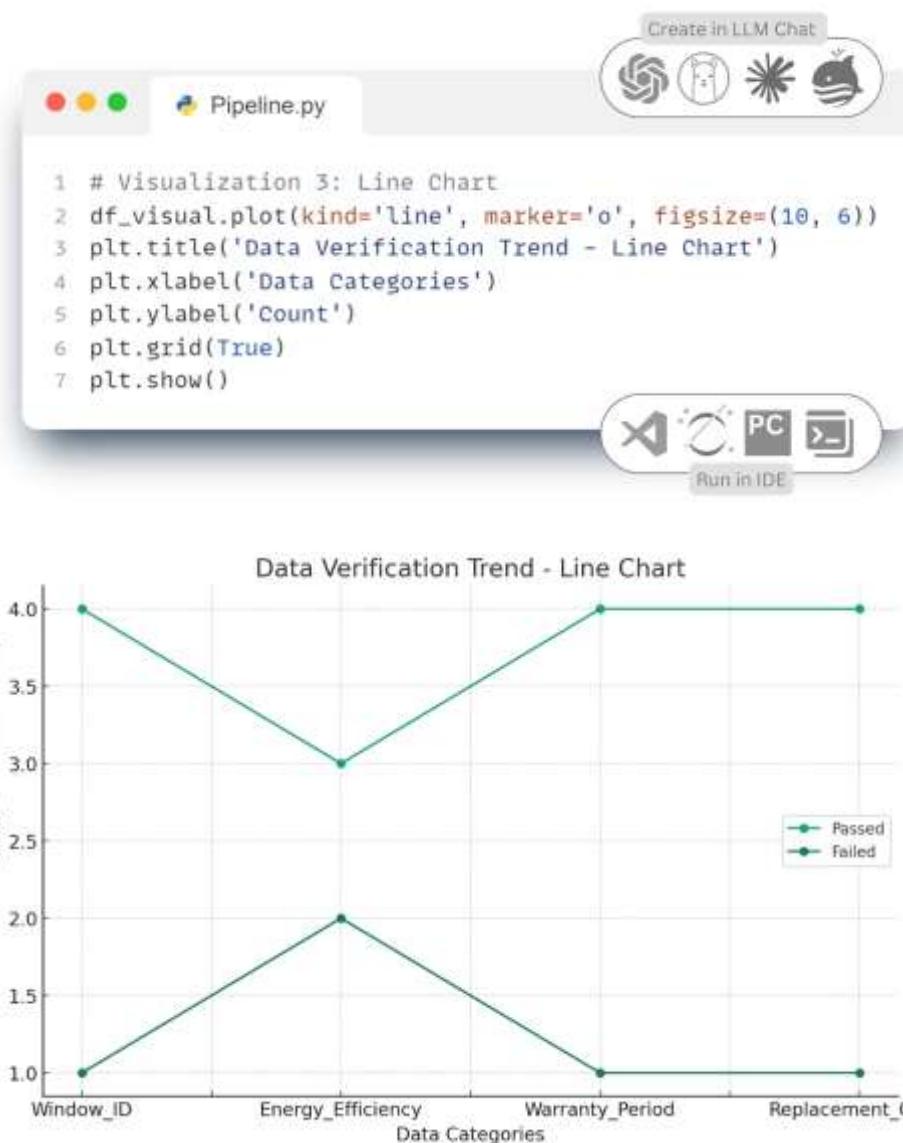
```

Фиг. 7.2-9 Визуализация на резултатите от стъпката Transform за проверка на стойностите на атрибутите от системата RPM (фиг. 7.2-7) като хистограма в стъпката Load .

- Съществуват много бесплатни библиотеки за визуализация с отворен код, които ви позволяват да представяте структурирани данни в различни формати. Нека да продължим да визуализираме резултатите с друг вид графика със следната подказа в чата:

Построяване на същите данни като графика от линии ↴

- Отговор на LLM :



Фигура 7.2-10 Визуализация на данните от валидирането (Фигура 7.2-8) като линейна диаграма, получена с помощта на библиотеката Matplotlib .

Съществуват много бесплатни библиотеки за визуализация с отворен код, като например:

- Seaborn - за статистически графики (фиг. 7.2-11)
- Plotly - за интерактивни уеб визуализации (фиг. 7.2-12, фиг. 7.1-6)
- Altair - за декларативна визуализация
- Dash или Streamlit - за създаване на пълни информационни табла

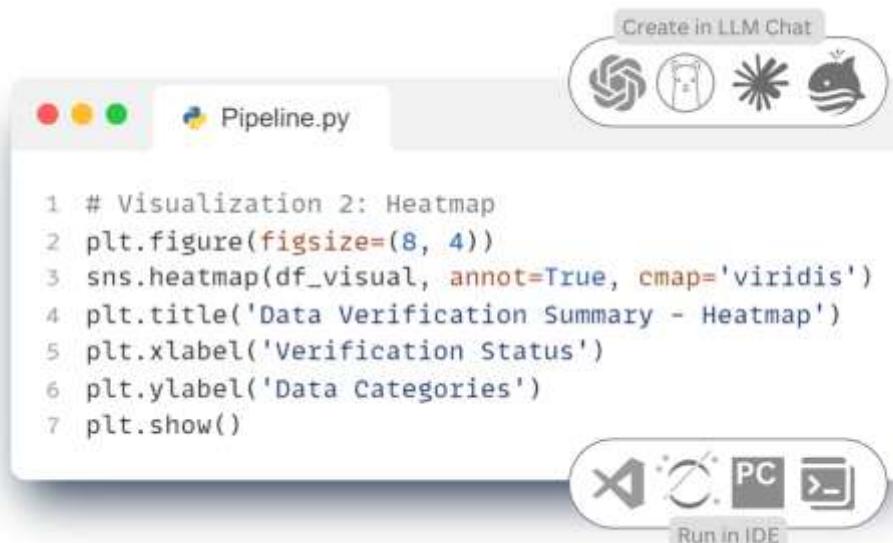
Не е необходимо да се познават конкретни библиотеки за визуализация - съвременните инструменти, включително LLM, позволяват автоматично генериране на код за графики и цели приложения въз основа на описанието на задачата.

Изборът на инструмент зависи от целите на проекта: дали става въпрос за доклад, презентация или онлайн табло. Например библиотеката с отворен код Seaborn е особено добра за работа с категорични данни, като помага да се идентифицират модели и тенденции.

- За да видите работещата библиотека Seaborn на адрес, можете да поискате директно от МЛНЗ да използва желаната библиотека или да из pratите подобно текстово запитване до МЛНЗ на адрес:

Показване на топлинна карта за ↗

- Отговорът на LLM е под формата на код и готова графика, чийто код за изчертаване вече може да се копира в IDE, а самата графика може да се копира или запази за поставяне в документ:



The screenshot shows the Pipeline.py application interface. At the top, there are three colored window control buttons (red, yellow, green) and a title bar with the logo and the text "Pipeline.py". Above the code editor, there is a button labeled "Create in LLM Chat". Below the code editor, there is a "Run in IDE" button with icons for VS Code, Docker, and PC. The code editor contains the following Python code:

```

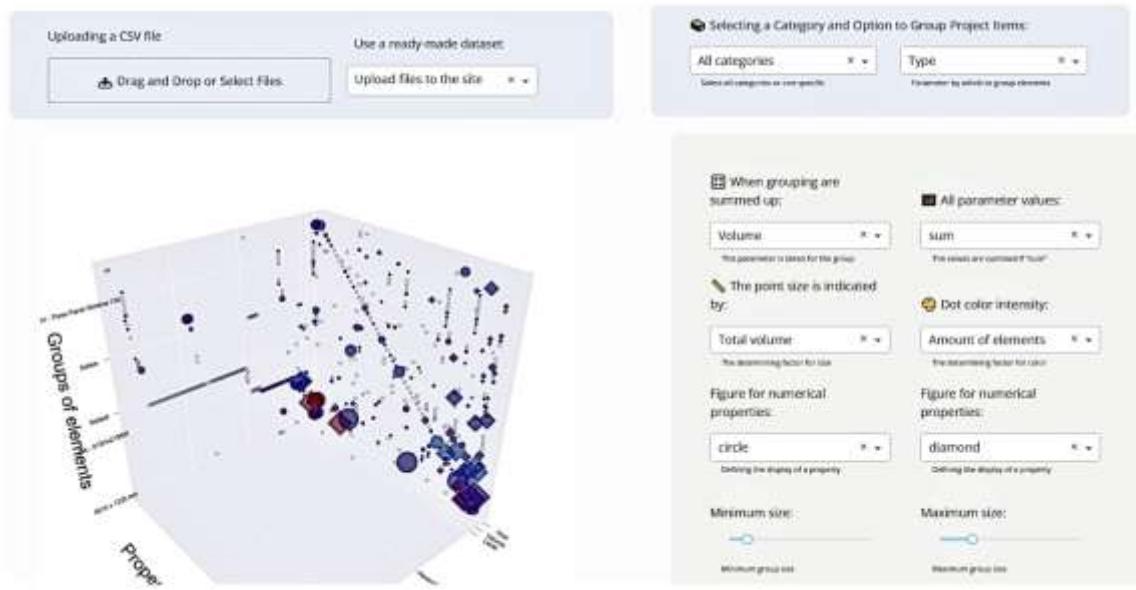
1 # Visualization 2: Heatmap
2 plt.figure(figsize=(8, 4))
3 sns.heatmap(df_visual, annot=True, cmap='viridis')
4 plt.title('Data Verification Summary - Heatmap')
5 plt.xlabel('Verification Status')
6 plt.ylabel('Data Categories')
7 plt.show()

```



**Фигура 7.2-11** Визуализиране на резултатите от валидирането (Фигура 7.2-8) на данните с помощта на библиотеката Seaborn.

За тези, които предпочитат интерактивен подход, има инструменти, които ви позволяват да създавате динамични диаграми и панели с възможност за взаимодействие. Библиотеката Plotly (Фигура 7.1-6, Фигура 7.2-12) предлага възможност за създаване на силно интерактивни диаграми и панели, които могат да бъдат вградени в уеб страници и позволяват на потребителя да взаимодейства с данните в реално време.



**Фиг. 7.2-12** Интерактивна 3D-визуализация на атрибути на елементи от CAD- (BIM-) проект с помощта на библиотеката Plotly.

Специализираните библиотеки с отворен код Bokeh, Dash и Streamlit предоставят удобен начин за представяне на данни, без да се изискват задълбочени познания по уеб разработване. Bokeh е подходяща за сложни интерактивни графики, Dash се използва за изграждане на пълноценни аналитични табла, а Streamlit ви позволява бързо да създавате уеб приложения за анализ на данни.

С помощта на такива инструменти за визуализация разработчиците и анализаторите могат ефективно да разпространяват резултатите сред колеги и заинтересовани страни, като позволяват интуитивно взаимодействие с данните и опростяват вземането на решения.

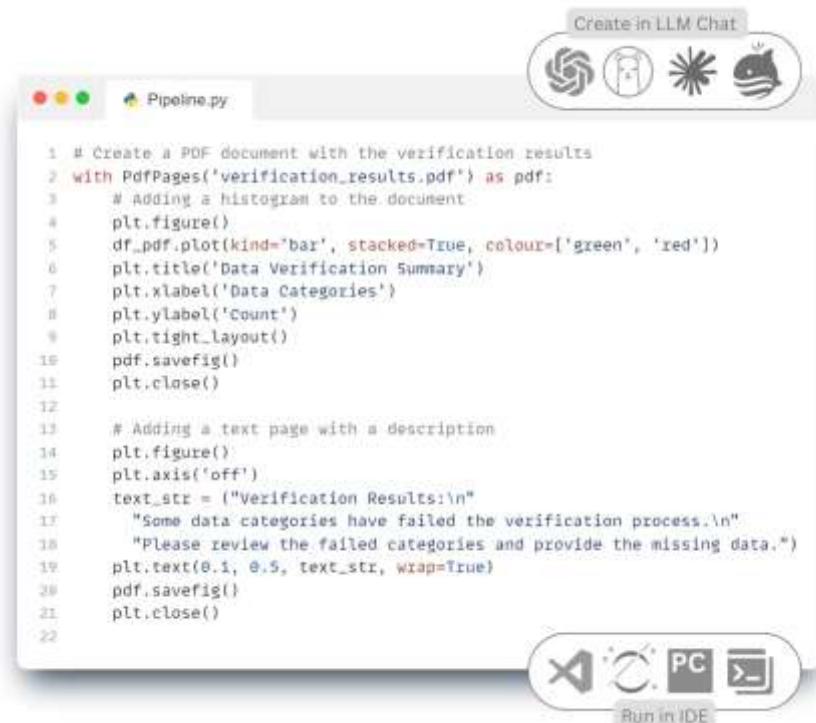
## ETL Зареждане: Автоматично създаване на PDF документи

На етапа на зареждане на данните можете не само да визуализирате данните, да ги качвате в таблици или бази данни, но и автоматично да генерирате отчети, включващи необходимите графики, диаграми и ключови аналитични показатели, които да бъдат получени от мениджъра или специалиста, очакващ резултатите от одита. Автоматизираните отчети могат да съдържат както коментари, така и текстова интерпретация на данните и елементи за визуализация - таблици, графики.

- За да създадете PDF -доклад с хистограма (фиг. 7.2-9) и описание на анализа въз основа на валидирането, което извършихме в предишните глави, е достатъчно да формулирате заявка в продължение на диалога с LLM, например:

Напишете код за създаване на PDF -файл с хистограма и описание на резултатите от валидирането на данните по-горе (в чата) и напишете текстово предупреждение, че някои категории не са били валидирани и че липсващите трябва да бъдат попълнени. ↴

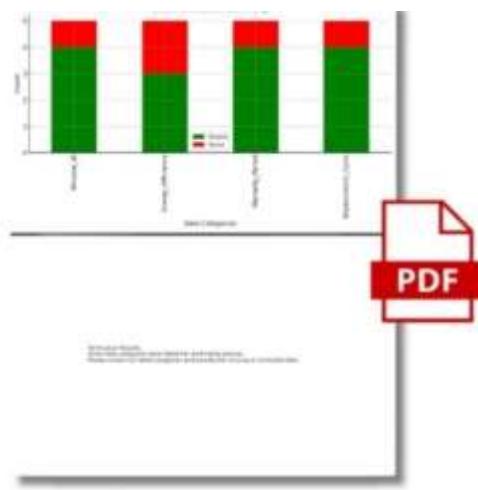
- Отговорът на LLM като код и завършен PDF с резултати:



```

1 # Create a PDF document with the verification results
2 with PdfPages('verification_results.pdf') as pdf:
3     # adding a histogram to the document
4     plt.figure()
5     df_pdf.plot(kind='bar', stacked=True, colour=['green', 'red'])
6     plt.title('Data Verification Summary')
7     plt.xlabel('Data Categories')
8     plt.ylabel('Count')
9     plt.tight_layout()
10    pdf.savefig()
11    plt.close()
12
13    # Adding a text page with a description
14    plt.figure()
15    plt.axis('off')
16    text_str = ("Verification Results:\n"
17                "Some data categories have failed the verification process.\n"
18                "please review the failed categories and provide the missing data.")
19    plt.text(0.1, 0.5, text_str, wrap=True)
20    pdf.savefig()
21    plt.close()
22

```



**Фигура 7.2-13 Автоматизираният код създава PDF -документ, съдържащ хистограма с тестови данни и текст с резултатите от теста.**

Автоматично написаното решение от само 20 реда код с помощта на LLM независимо създава желания PDF (или DOC) документ с визуализация под формата на хистограма на атрибутите (фиг. 7.2-13), показваща броя на данните, които са преминали и не са преминали валидирането, и с добавен текстов блок, обобщаващ резултатите и препоръките по-нататъшни действия.

Автоматизираното генериране на документи е ключов елемент от етапа на натоварване, особено в проектна среда, където бързината на отчитане и точността са от решаващо значение.

## ETL Load: автоматично генериране на документи от FPDF

Автоматизиране на отчитането на етапа ETL Load е важна стъпка в обработката на данни, особено когато резултатите от анализите трябва да бъдат представени във формат, който е лесен за предаване и разбиране. В строителния бранш това често е от значение за доклади за напредъка, статистически данни за проекта, доклади за осигуряване на качеството или финансова документация.

Един от най-удобните инструменти за такива задачи е библиотеката с отворен код, FPDF, достъпна както за Python, така и за PHP.

Библиотеката с отворен код **FPDF** предоставя гъвкав начин за генериране на документи чрез код, като ви позволява да добавяте заглавия, текст, таблици и изображения. Използването на код вместо ръчно редактиране намалява грешките и ускорява процеса на изготвяне на отчети в PDF формат.

Една от основните стъпки при създаването на PDF -документ е добавянето на заглавия и основен текст под формата на коментари или описания. При създаването на отчет обаче е важно не само да се добави текст, но и той да се структурира правилно. Заглавия, отстъпи, разстояние между редовете - всичко това влияе върху четимостта на документа. С помощта на FPDF, можете да задавате параметри за форматиране, да контролирате подредбата на елементите и да персонализирате стила на документа.

FPDF по принцип е много подобен на HTML. Тези, които вече са запознати с HTML, могат лесно да генерират PDF документи с всякаква сложност с помощта на FPDF, тъй като структурата на кода е много подобна на HTML маркировката: заглавия, текст, изображения и таблици се добавят по подобен начин. Тези, които не са запознати с HTML, не трябва да се притесняват - можете да използвате LLM, който веднага ще ви помогне да съставите кода, за да генерирате желаното оформление на документа.

- Следващият пример демонстрира как да генерирате отчет със заглавие и основен текст. Изпълнението на този код във всеки IDE с поддръжка на Python създава PDF - файл, съдържащ желаните заглавие и текст:

```
from fpdf import FPDF      # Импортиране на библиотеката FPDF
pdf = FPDF()    # Създаване на PDF -document
pdf.add_page()   # Добавяне на страница

pdf.set_font("Arial", style='B', size=16) # Задаване на шрифт: Arial, bold, size 16
pdf.cell(200, 10, "Отчет за проекта", ln=True, align='C') # Създайте заглавие и го центрирайте
pdf.set_font("Arial", size=12) # Променете шрифта на обикновен Arial, размер 12
pdf.multi_cell(0, 10, "Този документ съдържа данни за резултатите от проверката на проектното досие...") # Добавяне на многоредов текст
pdf.output(r "C:\reports\report.pdf") # Запазване на PDF -file
```



**Фигура 7.2-14 С няколко реда код на Python можем автоматично да генерираме необходимия ни PDF текстов документ.**

При изготвянето на доклади е важно да се вземе предвид, че данните, от които се формира документът, рядко са статични. Заглавията, текстовите блокове (фиг. 7.2-14) често се формират динамично, като получават стойности на етапа Transform в процеса ETL.

Използването на кода ви позволява да създавате документи, които съдържат актуална информация: име на проекта, дата на генериране на отчета, както и информация за участниците или текущото състояние. Използването на променливи в кода ви позволява автоматично да вмъкнете тези данни на необходимите места в отчета, като напълно елиминира необходимостта от ръчно редактиране преди изпращане.

В допълнение към обикновения текст и заглавията таблиците заемат специално място в проектната документация. Почти всеки документ съдържа структурирани данни: от описания на обекти

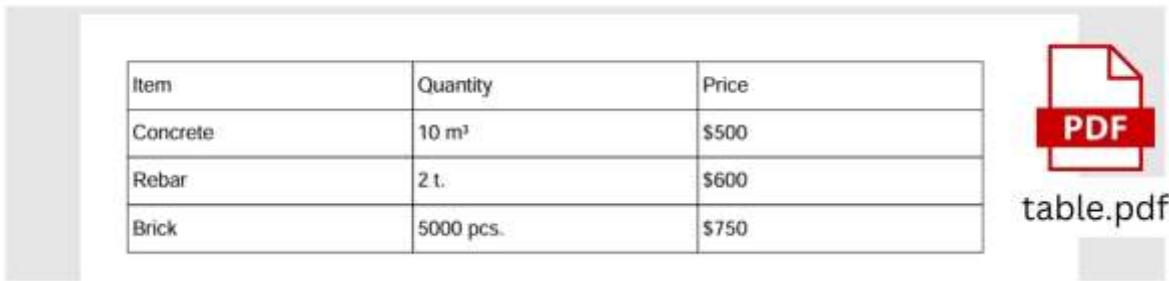
до резултати от проверки. Автоматичното генериране на таблици въз основа на данни от етапа на трансформиране позволява не само да се ускори процесът на подготовка на документите, но и да се сведат до минимум грешките при прехвърляне на информацията. FPDF позволява да се вмъкват таблици в PDF -файлове (като текст или картинки), като се задават граници на клетките, размери на колоните и шрифтове (фиг. 7.2-15). Това е особено удобно при работа с динамични данни, когато броят на редовете и колоните може да се променя в зависимост от задачите на документа.

- Следващият пример показва как да се автоматизира създаването на таблици, например със списъци на материали, оценки или резултати от изпитвания на параметри:

```
дани = [
    ["Артикул", "Количество", "Цена"], # Заглавия на колони
    ["Бетон", "10 м3", "$ 500."], # Данни за първия ред
    ["Арматура", "2 тона", "$ 600"], # Данни от втория ред.
    ["Тухли", "5 000 броя", "$ 750."], # Данни от трети ред.
]

pdf = FPDF () # Създаване на PDF -document
pdf.add_page() # Добавяне на страница
pdf.set_font("Arial", size=12) # Задаване на шрифт

for row in data: # Търсене на редове в таблицата
    for item in row: # Преминете през клетките в реда
        pdf.cell(60, 10, item, border=1) # Създаване на клетка с рамка, широчина 60 и височина
10
pdf.ln() # Преминаване към следващия ред
pdf.output(r "C:\reports\table.pdf") # Запазване на PDF -file
```



Item	Quantity	Price
Concrete	10 m <sup>3</sup>	\$500
Rebar	2 t.	\$600
Brick	5000 pcs.	\$750

**Фигура 7.2-15** Можете автоматично да генерирате не само текст, но и всякаква информация за таблици от стъпката Transform в PDF.

В реалните сценарии на отчитане табличите обикновено са динамично генерирана информация, получена на етапа на преобразуване на данните. В дадения пример (фиг. 7.2-15) таблицата е вмъкната в PDF -документ в статичен вид: данните за примера са поставени в речника на данните (първи ред на кода), в реални условия такава променлива се попълва автоматично след напр. групиране на рамки от данни.

На практика такива таблици често се изграждат въз основа на структурирани данни, идващи от

различни динамични източници: бази данни, Excel -файлове, API -интерфейси или резултати от аналитични изчисления. Най-често на етапа на трансформиране (ETL) данните се обобщават, групират или филтрират - и едва след това се трансформират в общи суми под формата на графики или двумерни таблици, показвани в отчети. Това означава, че съдържанието на таблиците може да се променя в зависимост от избраните параметри, периода на анализ, филтрите на проекта или потребителските настройки.

Използването на динамични рамки за данни и набори от данни в етапа Transform прави процеса на отчитане в етапа Load възможно най-гъвкав, мащабираме и лесно повтаряме, без да е необходима ръчна намеса.

В допълнение към таблиците и текста FPDF поддържа и добавяне на графики на таблични данни, което ви позволява да вграждате в отчета изображения, генерирали с Matplotlib или други библиотеки за визуализация, които разглеждахме по-горе. Всякакви графики, диаграми и диаграми могат да се добавят към документа с помощта на кода.

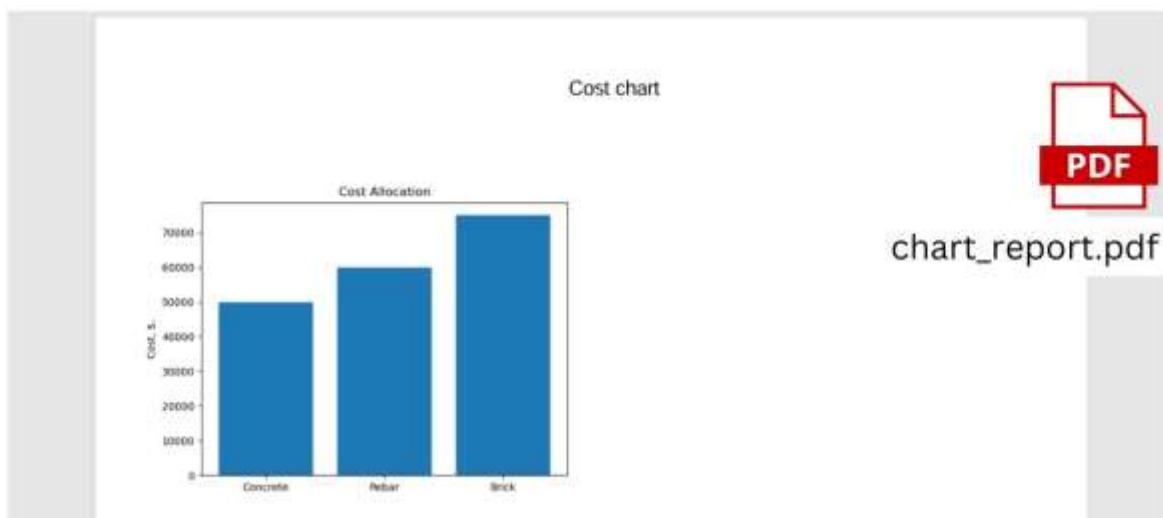
- Използвайки библиотеката на Python FPDF, нека добавим към PDF документа графика, предварително генерирана с Matplotlib:

```
import matplotlib.pyplot as plt # Импортиране на Matplotlib за създаване на графики

fig, ax = plt.subplots() # Създаване на фигурата и осите на графиката
categories = ["Concrete", "Rebar", "Brick"] # Имената на категориите
values = [50000, 60000, 75000] # Стойности на категориите
ax.bar(categories, values) # Създаване на стълбовидна диаграма
plt.ylabel("Стойност,$.") # Подпишете оста Y
plt.title("Разпределение на разходите") # Добавете заглавие
plt.savefig(r "C:\reports\chart\chart\chart.png") # Запазване на диаграмата като изображение

pdf = FPDF () # Създаване на PDF-document
pdf.add_page() # Добавяне на страница
pdf.set_font("Arial", size=12) # Задаване на шрифт
pdf.cell(200, 10, "Cost Chart", ln=True, align="C") # Добавяне на заглавие

pdf.image(r "C:\reports\chart\chart\chart.png", x=10, y=30, w=100) # Вмъкване на изображение в PDF (x, y - координати, w - ширина)
pdf.output(r "C:\reports\chart_report.pdf") # Запазване на PDF файла
```



**Фиг. 7.2-16 С десетина реда код можете да генерирате графика, да я запазите и след това да я поставите в PDF документ.**

С FPDF процесът на подготовка на документи и логика става прозрачен, бърз и удобен. Шаблоните, вградени в кода, позволяват генерирането на документи с актуални данни, като отпада необходимостта от ръчно попълване.

Използване на автоматизацията на ETL - вместо отнемащите време ръчни отчети, специалистите могат да се съсредоточат върху анализа на данните и вземането на решения, а не върху избора на правилния инструмент за работа с определен силоз от данни с ясен потребителски интерфейс.

По този начин библиотеката FPDF предоставя гъвкав инструмент за автоматизирано създаване на документи с всякаква сложност - от кратки технически доклади до сложни аналитични обобщения с таблици и диаграми, което позволява не само да се ускори документооборотът, но и значително да се намали вероятността от грешки, свързани с ръчното въвеждане и форматиране на данни.

## ETL Натоварване: Отчитане и зареждане на други системи

На етапа Load резултатите бяха генериирани под формата на таблици, графики и окончателни доклади в PDF формат, изгответи в съответствие с установените изисквания. Освен това е възможно тези данни да се експортират в машинночитаеми формати (напр. CSV), което е необходимо за интегриране с външни системи като ERP, CAFM, CPM, BI платформи и други корпоративни или индустриални решения. В допълнение към CSV, качването може да се извърши в XLSX, JSON, XML или директно в бази данни, които поддържат автоматичен обмен на информация.

- 💡 За да генерирате подходящ код за автоматизиране на стъпката Load, просто направете заявка към интерфейса LLM, например: ChatGPT, LlaMa, Mistral DeepSeek, Grok, Claude или QWEN:

Напишете код за генериране на отчет за резултатите от валидирането на данните в DataFrame, където колоните с префикс "verified\_" се преброяват, преименуват се на "Passed" (преминал) и "Failed" (не преминал), липсващите стойности се заменят с 0, след което само тези редове, които са преминали всички валидирания, се експортират в CSV -файл.

 Отговор на LLM:



The screenshot shows a code editor window titled "Pipeline.py". The code is as follows:

```

1 # Create a report on the results of data validation
2 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
3 report.columns = ['Passed', 'Failed']
4 report.fillna(0, inplace=True)
5
6 # Export the cleaned and verified data to a CSV file
7 verified_df = df[df.filter(like='verified_').all(axis=1)]
8 verified_df.to_csv('verified_data.csv', index=False)

```

At the top of the window, there are icons for "Create in LLM Chat" and four AI models. At the bottom, there are icons for "Run in IDE", "PC", and "Run in browser".

Фигура 7.2-17 Утвърдените данни, получени в етапа на преобразуване от крайната рамка от данни, се експортират в CSV -файл за интегриране с други системи.

Горният код (фиг. 7.2-17) реализира последния етап на ETL -процеса - Load (Зареждане), по време на който проверените данни се записват във формат CSV, съвместим с повечето външни системи и бази данни. По този начин завършихме пълния цикъл на ETL -процеса, включващ извлечение, трансформиране, визуализиране, документиране и експортиране на данни към необходимите ни системи и формати, което осигурява възпроизвеждаемост, прозрачност и автоматизация на работата с информация.

Конвейерът ETL може да се използва както за обработка на единични проекти, така и за широкомащабни приложения - когато се анализират стотици или хиляди входящи данни под формата на документи, изображения, сканирания, CAD проекти, облаци от точки, PDF файлове или други източници, идващи от разпределени системи. Възможността за пълно автоматизиране на процеса превръща ETL не само в технически инструмент за обработка, но и в основа на цифрова строителна информационна инфраструктура.

## ETL с LLM: Визуализиране на данни от PDF -документи

Време е да преминете към изграждане на пълноценен процес ETL, който обхваща всички ключови етапи на обработка на данни в един сценарий - извлечение, преобразуване и зареждане. Нека да изградим автоматизиран ETL-пилот, който ви позволява да обработвате PDF документи без ръчна

работка - извличане на данни от документи, визуализиране, анализ и прехвърляне към други системи.

Процесът ETL в нашия пример ще бъде описан чрез подкани, които ще трябва да обяснят на езиковия модел (LLM) всички процеси ETL с описание на крайния резултат, който трябва да се получи. В този случай задачата е да се намерят всички PDF файлове в посочената папка и нейните подпапки, да се извлече съответната информация от тях - напр. имена на материали, количества и разходи - и да се представи резултатът като структурирана таблица (DataFrame) за по-нататъшен анализ

- 👉 Първа текстова заявка в LLM за автоматично извличане на данни от множество PDF документи и създаване на рамка от данни за стъпката "Извличане":

Напишете код за извличане на информация за материала от PDF файлове в дадена папка и нейните подпапки. Данните в PDF файла включват име на материала, количество и цена. Резултатът трябва да се запише в DataFrame ↵

💡 Отговор на LLM :



The screenshot shows a LLM interface with a code editor window titled "Pipeline.py". The code is a Python script for extracting data from PDF files. It uses the pdfplumber library to process each page and extract text. A specific function, "extract\_data", is defined to walk through a base folder and its subfolders, processing PDF files. The extracted text is split into lines, and if a line contains the word "Material", it is split into parts and converted into a dictionary with keys "file", "material", "quantity", and "cost". These dictionaries are then converted into a pandas DataFrame. The code also includes a comment specifying the base folder path as "C:\\".

```

1 import os
2 import pdfplumber
3 import pandas as pd
4
5 def extract_data(base_folder_path):
6     """
7         Extract material data from PDF files in the specified base
8         folder and all its subfolders. Assumes each line in the PDF contains
9         'Material' followed by material name, quantity, and cost.
10    """
11    data = []
12    # Walk through each folder and subfolder in the base folder
13    for folder_path, _, files in os.walk(base_folder_path):
14        for file in filter(lambda f: f.endswith('.pdf'), files):
15            with pdfplumber.open(os.path.join(folder_path, file)) as pdf:
16                # Process each page in the PDF
17                for page in pdf.pages:
18                    # Splitting text into lines
19                    for line in page.extract_text().split('\n'):
20                        # Check if the line contains material data
21                        if 'Material' in line:
22                            parts = line.split()
23                            # Append data to list; convert quantity and cost to float
24                            data.append({'file': file, 'material': parts[1],
25                                         'quantity': float(parts[2]), 'cost': float(parts[3])})
26    # Convert list of dictionaries to pandas DataFrame
27    return pd.DataFrame(data)
28
29 # Specify the base folder path containing PDF files and subfolders with PDFs
30 df = extract_data(r'C:\')

```

[Create in LLM Chat](#)

[Run in IDE](#)

**Фигура 7.2-18 LLM създава Python -код за извлечане на данни от PDF -файлове в определена папка и всички нейни подпапки.**

Отговорът на LLM (Фигура 7.2-18) е готов Python -скрипт, който автоматично обхожда всички папки, отваря намерените PDF -файлове, извлича текстова информация от тях и я преобразува в таблица. Полученият чат-код може да бъде стартиран в някоя от популярните IDE PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с PyDev плъгин, Thonny, Wing IDE, IntelliJ IDEA с Python плъгин, JupyterLab или популярните онлайн инструменти Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Алтернативно, за да опростим процеса на събиране, вместо да копираме кода от LLM и да използваме кода в IDE, можем също така да качим десетина PDF -файла директно в чата на LLM (Фигура 7.2-19) и да получим таблица като изход, без да се налага да виждаме кода или да го стартираме. Резултатът от изпълнението на този код ще бъде таблица с атрибутите, които сме избрали.

ChatGPT 4 -

You  
Write code to extract material information from PDF files in a given folder. The data in the PDF includes material name, quantity and cost. The result should be saved in a pandas DataFrame and displayed on the screen.

ChatGPT  
Below is a sample Python script that demonstrates this process using 'PyPDF2'. Please note that you might need to adjust the text parsing part based on the specific format of your PDF files.

```
python
import os
import PyPDF2
import pandas as pd

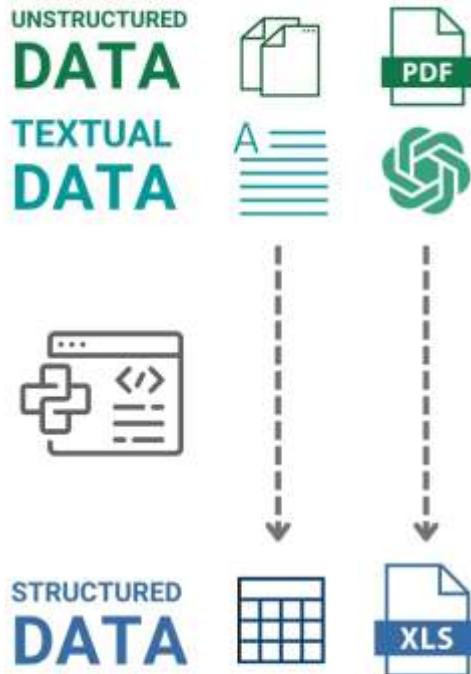
# Create a DataFrame from the extracted data
df = pd.DataFrame(data)

# Display the DataFrame
print(df)
```

Here's an example of how the output would look in a pandas DataFrame after extracting and processing the material information from the PDF files:

	Material Name	Quantity	Cost
0	Copper Wire	100	250.50
1	Aluminum Sheet	50	150.75
2	Plastic Pipes	200	100.00
3	Steel Rods	80	300.20

This table shows the material name, quantity, and cost as columns, with each row representing the data extracted from a PDF file. The actual content and structure of your DataFrame will depend on the specific format and data in your PDF files. [-]

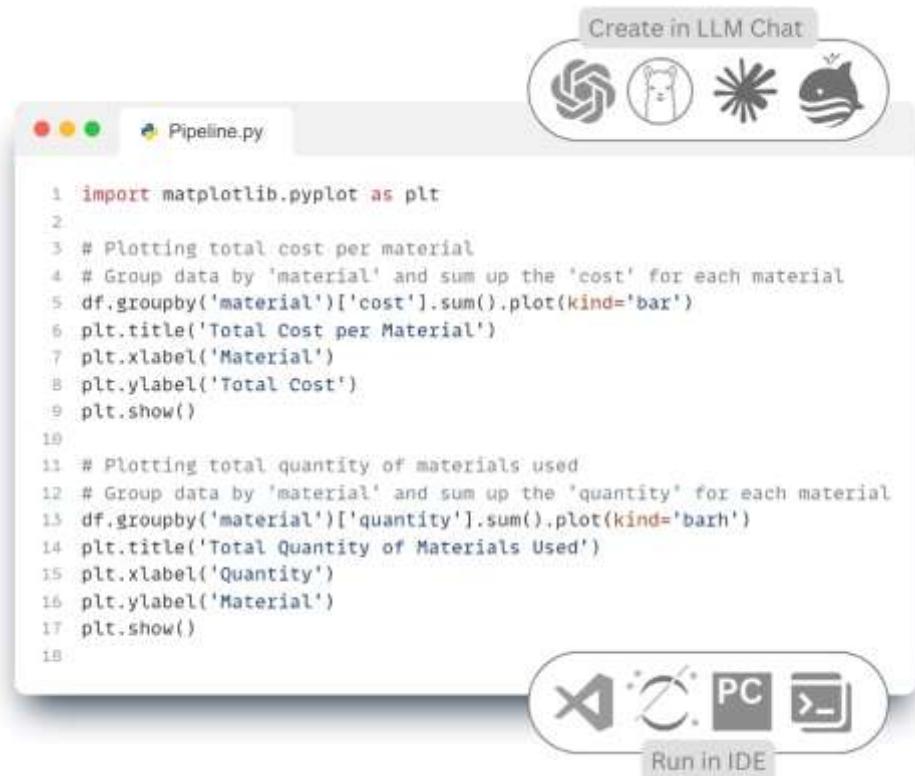


Фигура 7.2-19 Резултатът от изпълнението на код в LLM, който извлича данни от PDF - файлове в структуриран изглед на рамка от данни с избрани атрибути.

На следващия етап искаме езиков модел на получените данни - например за сравняване на разходите и обема на използване на материалите и създаване на някои примерни визуализации, които ще послужат като основа за по-нататъшен анализ.

- Поискайте в продължителен разговор с LLM да начертаете някои графики от таблиците, които бяха изгответи в етапа на преобразуване (фигура 7.2-18):

Визуализирайте общите разходи и количеството на всеки материал от DataFrame (Фигура 7.2-18)



The screenshot shows a Python code editor window titled "Pipeline.py". The code is as follows:

```

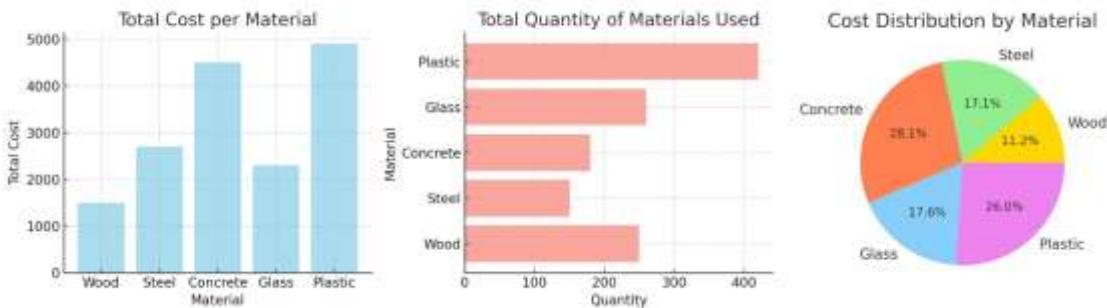
1 import matplotlib.pyplot as plt
2
3 # Plotting total cost per material
4 # Group data by 'material' and sum up the 'cost' for each material
5 df.groupby('material')['cost'].sum().plot(kind='bar')
6 plt.title('Total Cost per Material')
7 plt.xlabel('Material')
8 plt.ylabel('Total Cost')
9 plt.show()
10
11 # Plotting total quantity of materials used
12 # Group data by 'material' and sum up the 'quantity' for each material
13 df.groupby('material')['quantity'].sum().plot(kind='barh')
14 plt.title('Total Quantity of Materials Used')
15 plt.xlabel('Quantity')
16 plt.ylabel('Material')
17 plt.show()
18

```

At the top right of the window are icons for "Create in LLM Chat" and four AI models. At the bottom right is a "Run in IDE" button.

**Фигура 7.2-20 LLM отговор -модели като код на Python за визуализиране на данни от рамка с данни с помощта на библиотеката matplotlib.**

LLM автоматично генерира и изпълнява Python -код (фиг. 7.2-20), използвайки библиотеката matplotlib. След изпълнението на този код получаваме графики на разходите и използването на материали в строителните проекти директно в чата (фиг. 7.2-21), което значително опростява аналитичната работа.

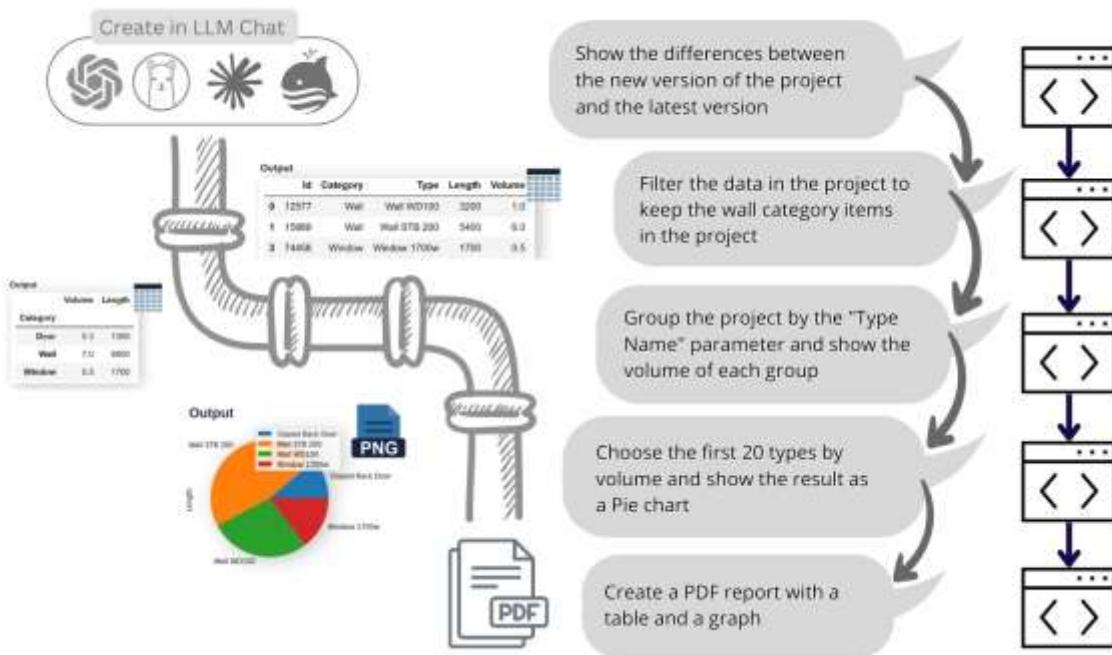


**Фигура 7.2-21 Визуализация на отговора на LLM като графики въз основа на данните, събрани в DataFrame.**

Подкрепата за разработване на идеи за писане на код за ETL, анализиране и изпълнение на кода и визуализиране на резултатите е достъпна чрез прости текстови заявки в LLM, без да е необходимо да се изучават основите на програмирането. Появата на инструменти за изкуствен интелект като LLM определено променя подхода към програмирането и автоматизирането на обработката на

данни (фигура 7.2-22).

Според доклада на PwC "Каква е реалната стойност на изкуствения интелект за вашия бизнес и как можете да се възползвате от него?" (2017 г.) [139], автоматизацията на процесите и подобряването на производителността ще бъдат основните двигатели на икономическия растеж. А подобренията в производителността се очаква да представляват повече от 55% от целия ръст на БВП, движен от изкуствения интелект, в периода 2017-2030 г."



Фигура 7.2-22 AI LLM помага да се генерира чернова на код, която се прилага в бъдещи проекти, без да е необходимо LLM.

С помощта на инструменти като ChatGPT, LLaMa, Mistral, Claude, DeepSeek, QWEN, Grok, както и на отворени данни и софтуер с отворен код, можем да автоматизираме процеси, които преди това са се извършвали само със специализирани, скъпи и трудни за поддръжка модулни патентовани системи.

В контекста на строителството това означава, че компаниите, които първи въведат автоматизираните процеси на тръбопроводите - данни, ще извлекат значителни ползи - от подобрена ефективност на управлението на проектите до намаляване на финансовите загуби и премахване на фрагментираните приложения и обособените складове за данни.

Описаната логика на изпълнение на бизнес задачите в процеса ETL е важна част от автоматизацията на процесите за анализ и обработка на данни, която е специфична разновидност на по-широко понятие - тръбопроводи (Pipelines).

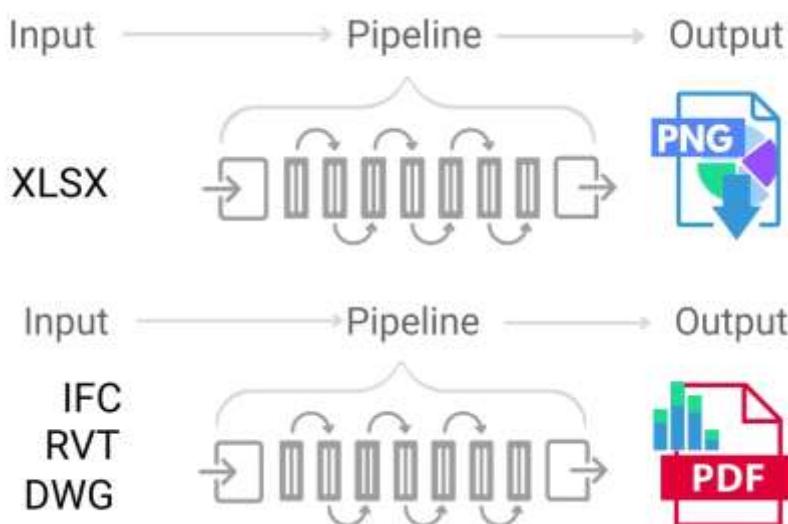
## ГЛАВА 7.3.

### АВТОМАТИЧЕН КОНВЕЙЕР ЗА ETL (ТРЪБОПРОВОД)

#### Тръбопровод: Автоматичен ETL конвейер данни

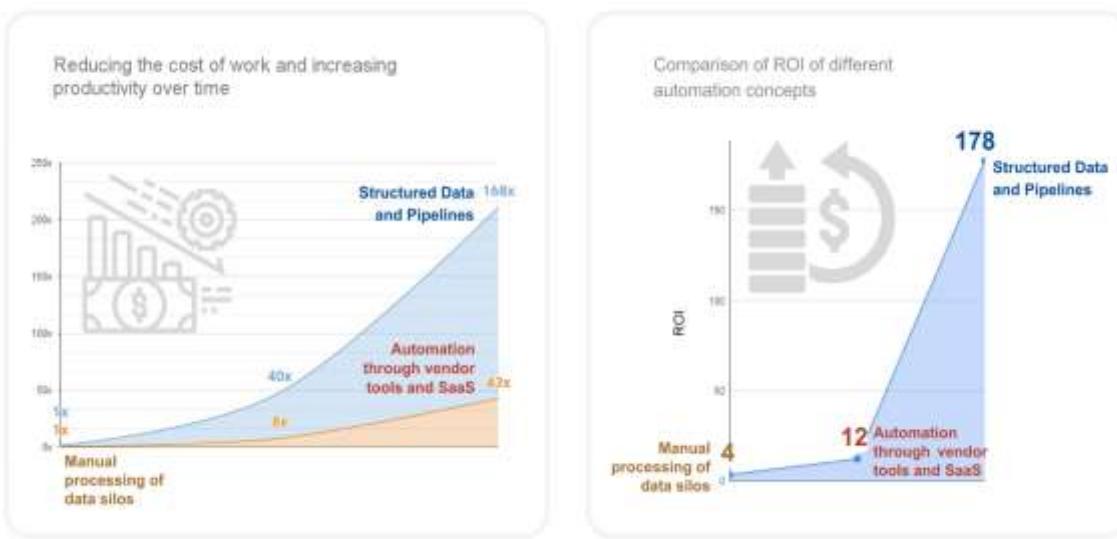
Процесът ETL традиционно се използва за обработка на данни в аналитични системи, като обхваща както структурирани, така и неструктурни източници. В днешната цифрова среда обаче все по-често се използва по-широк термин - Pipeline (конвейер), който описва всяка последователна верига на обработка, при която изходът от един етап става вход за следващия.

Този подход се прилага не само за данните, но и за други видове автоматизация: обработка на задачи, изграждане на отчети, интеграция със софтуер и цифров работен процес (фиг. 7.3-1).



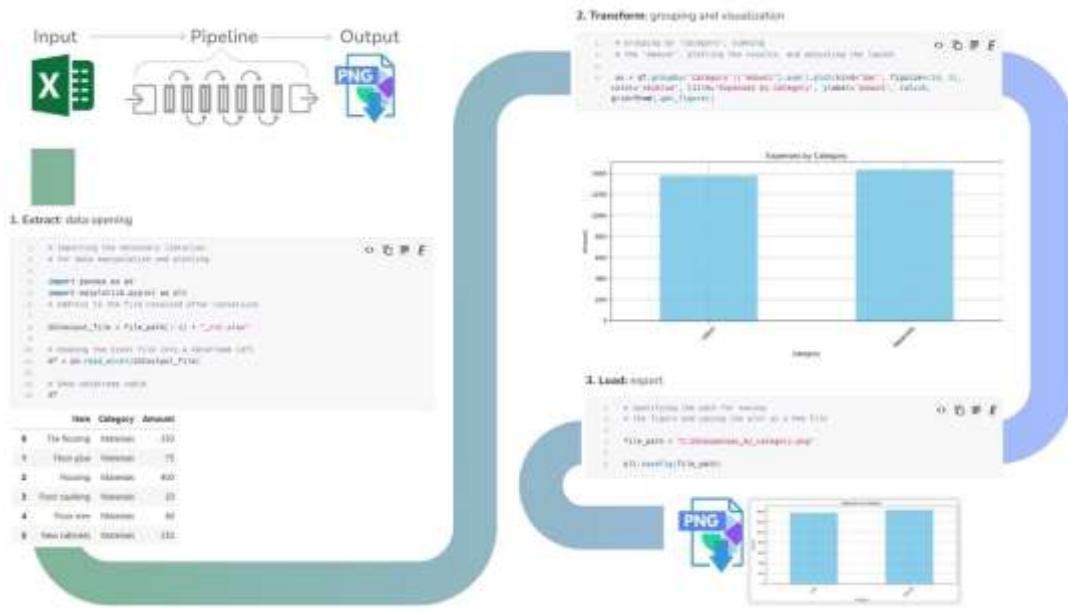
Фигура 7.3-1 Тръбопровод е последователност на обработка, при която изходът на един етап става вход за следващия етап.

Използването на тръбопроводи е един от основните елементи на автоматизацията, особено при работа с големи количества разнородни данни. Архитектурата на тръбопроводите позволява сложните стъпки за обработка да бъдат организирани в модулен, последователен и управляем формат, което увеличава четимостта, опростява поддръжката на кода и дава възможност за поетапно отстраняване на грешки и мащабиране на тестуване.



**Фиг. 7.3-2 ROI** Процесът на валидиране на данни по тръбопровод намалява времето за изпълнение с десетки и стотици пъти в сравнение с обработката с класически инструменти [74].

За разлика от ръчната работа в собствени системи (ERP, PMIS, CAD и т.н.), конвейерът позволява значително (фиг. 7.3-2) да се увеличи скоростта на задачите, да се избегне повтарящата се работа и да се автоматизира стартирането на процесите в точното време (фиг. 7.3-3).

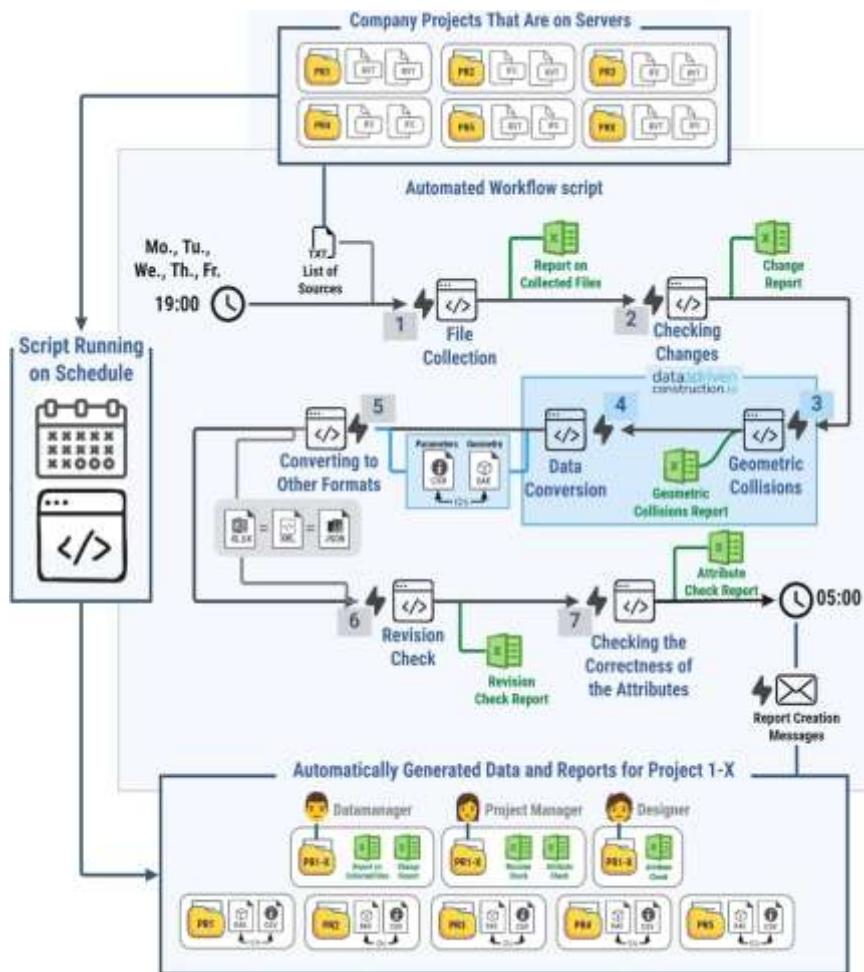


**Фиг. 7.3-3 Пример за ETL Конвейер за автоматично извлечане на диаграма от таблични данни в XLSX файл без отваряне на Excel.**

За да обработвате поточни данни и да изградите автоматизирана поточна линия, подобно на процеса ETL, трябва предварително да определите източниците на данни, както и времевата рамка за събирането им - за конкретен бизнес процес или за цялата компания.

В строителните проекти данните идват от много разнородни източници с различни интервали на актуализация. За да се създаде надеждна витрина с данни, е изключително важно да се записва кога се извлича и актуализира информацията. Това дава възможност за своевременно вземане на решения и подобрява ефективността на управлението на проекта.

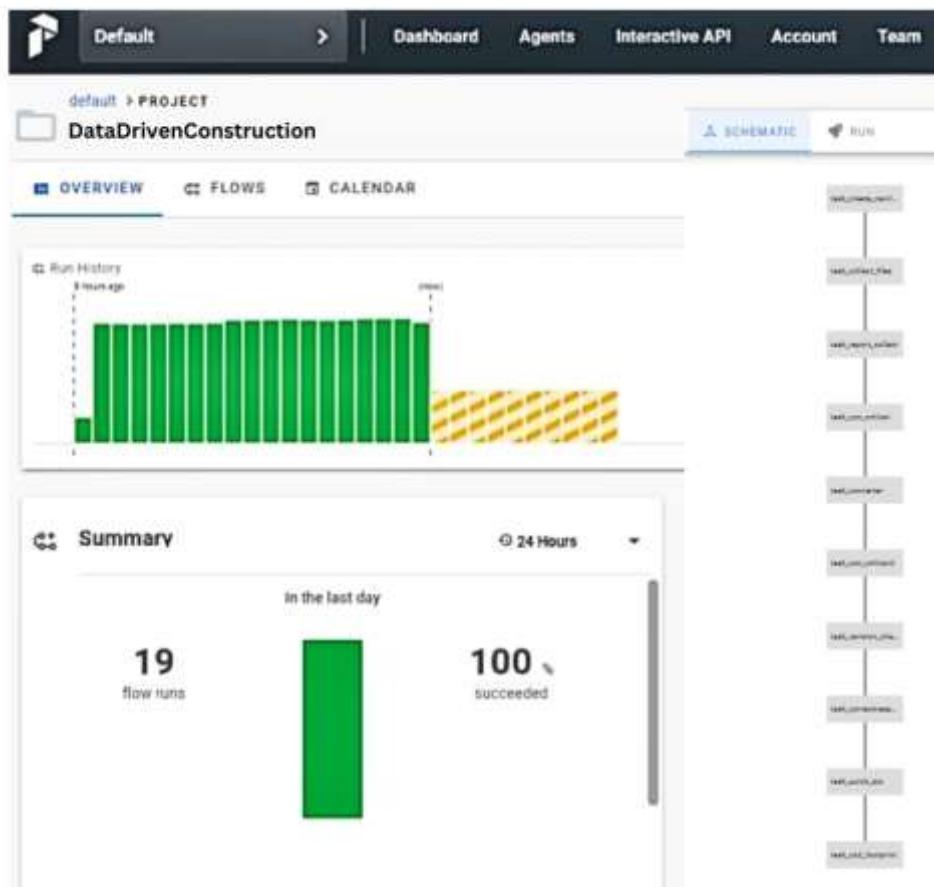
Един от вариантите е процесът на сглобяване да започне във фиксиран час - например в 19:00 ч., в края на работния ден. В този момент се активира първият скрипт, отговарящ за обобщаването на данните от различните системи и хранилища (фиг. 7.3-4 стъпка 1). Следва автоматична обработка и преобразуване на данните в структуриран формат, подходящ за анализ (фиг. 7.3-4 стъпка 2-4). На последния етап, като се използват подгответените данни, автоматично се генерират отчети, информационни табла и други продукти, описани в предходните глави (фиг. 7.3-4 стъпка 6-7). В резултат на това към 05:00 ч. сутринта мениджърите вече разполагат с актуални доклади за състоянието на проекта в необходимия формат (фиг. 7.3-5).



Фиг. 7.3-4 Данните в тръбопровода, събиращи автоматично вечерта, се обработват през нощта, така че на сутринта мениджърите да разполагат с актуални отчети и свежи доклади.

Навременното събиране на данни, определянето на ключови показатели за ефективност, автоматизацията на процесите на трансформация и визуализацията чрез информационни табла са ключови елементи на успешното вземане на решения, базирани на данни.

Такива автоматизирани процеси (фиг. 7.3-4) могат да се изпълняват напълно автономно: те се изпълняват по график, обработват данни без надзор и могат да бъдат разположени в облака или на собствен сървър на компанията (фиг. 7.3-5). Това позволява такива ETL конвейери да бъдат интегрирани в съществуващата ИТ инфраструктура, като се запазва контролът върху данните и се осигурява гъвкавост при мащабирането.



Фиг. 7.3-5 Автоматични ETL -конвейерни процеси (фиг. 7.3-4) на платформата Prefect, при които 10 python скрипта се изпълняват последователно след 19:00 часа всеки работен ден.

Автоматизирането на работните процеси не само повишава производителността на екипа, като освобождава време за по-смислени и по-малко рутинни задачи, но и служи като важна първа стъпка към включването на технологиите за изкуствен интелект (AI) в бизнес процесите, които ще разгледаме по-подробно в главата за прогнозен анализ и машинно обучение.

## Тръбопровод -ETL процес на валидиране на данни с LLM

В предишните глави за създаване на изисквания към данните и автоматизиране на ETL, стъпка по стъпка разбихме процеса на подготовка, преобразуване, валидиране и визуализация на данните. Тези дейности бяха реализирани като отделни блокове код (Фигура 7.2-18 - Фигура 7.2-20), всеки от които изпълнява конкретна задача.

Сега пред нас стои следващата цел - да обединим тези елементи в единен, съгласуван и автоматизиран конвейер на обработка на данни - конвейер, ETL - конвейер - в който всички етапи (зареждане, валидиране, визуализация, експортиране) се изпълняват последователно в един автоматично изпълним скрипт.

В следващия пример ще бъде реализиран пълен цикъл на обработка на данни: от зареждане на изходния CSV файл → до проверка на структурата и стойностите с помощта на регулярни изрази → изчисляване на резултатите → генериране на визуален отчет в PDF формат.

- Можете да използвате следната текстова заявка към LLM, за извлечане на съответния код:

Моля, напишете примерен код, който зарежда данни от CSV, валидира DataFrame с помощта на регулярни изрази, проверява идентификатори във формат "W-NEW" или "W-OLD", енергийна ефективност с букви от "A" до "G", гаранционен период и цикъл на подмяна с цифрови стойности в години и накрая създава отчет с брой на преминалите и непроменените стойности, генерира PDF с хистограма на резултатите и добавя текстово описание. ↴

■ Отговор на LLM :



```

1 # Importing necessary libraries
2 import pandas as pd
3 import re
4 import matplotlib.pyplot as plt
5 from matplotlib.backends.backend_pdf import PdfPages
6
7
8 # Load data
9 df = pd.read_csv('path_to_csv_file.csv')
10
11 # Define patterns and verify data
12 patterns = {
13     'Window_ID': r'^W-NEW|W-OLD[0-9]+|D-1[0-9]{2}$',
14     'Energy_Efficiency': r'^[A-G]\+*$',
15     'Warranty_Period': r'^\d+\syears$',
16     'Replacement_Cycle': r'^\d+\syears$'
17 }
18
19 for column, pattern in patterns.items():
20     df[f'verified_{column}'] = df[column].apply(lambda x: bool(re.match(pattern, str(x))))
21
22 # Create report
23 report = df.filter(like='verified_').apply(lambda x: x.value_counts()).T
24 report.columns = ['Passed', 'Failed']
25
26 # Generate PDF
27 pdf_path = '/mnt/data/verification_results_full_pipeline.pdf'
28 with PdfPages(pdf_path) as pdf:
29     # Plot and save bar chart
30     report.plot(kind='bar', stacked=True, color=['green', 'red'])
31     plt.title('Data Verification Summary')
32     plt.xlabel('Data Categories')
33     plt.ylabel('Count')
34     plt.tight_layout()
35     pdf.savefig()
36     plt.close()
37
38 # Add text page
39 plt.figure()
40 plt.axis('off')
41 text_str = "Verification Results Summary"
42 plt.text(0.1, 0.5, text_str, wrap=True)
43 pdf.savefig()
44 plt.close()
45
46 pdf_path
47

```

Фиг. 7.3-6 Pipeline (ETL) автоматизира пълния цикъл на обработка на данните: от изтеглянето и валидирането до създаването на структуриран отчет в PDF формат .

Автоматизираният код (фиг. 7.3-6) вътре в чата на LLM или в DIE, след като копирате кода, ще валидира данните от CSV -файла, като използва зададените регулярни изрази, ще създаде отчет за броя на преминалите и непреминалите записи и след това ще запише резултатите от валидирането като PDF -файл.

Тази структура на ETL - конвейер, при която всяка стъпка - от зареждането на данни до генерирането на отчети - е реализирана като отделен модул, осигурява прозрачност, мащабируемост и възпроизвеждаемост. Представянето на логиката за валидиране като лесен за четене код на Python прави процеса прозрачен и разбираем не само за разработчиците, но и за специалистите по управление на данни, качество и анализи.

Подходът на Pipeline към автоматизирането на обработката на данни ви позволява да стандартизираме процесите, да увеличим тяхната повторяемост и да опростим адаптирането към нови проекти. По този начин се създава единна методология за анализ на данни, независимо от източника или вида на задачата - независимо дали става въпрос за тестване на съответствието, генериране на отчети или прехвърляне на данни към външни системи.

Подобна автоматизация намалява човешките грешки, намалява зависимостта от патентовани решения и увеличава точността и надеждността на резултатите, което ги прави подходящи както за оперативни анализи на ниво проект, така и за стратегически анализи на ниво компания.

## **Тръбопровод -ETL: проверка на данните и информацията за елементите на проекта в CAD (BIM)**

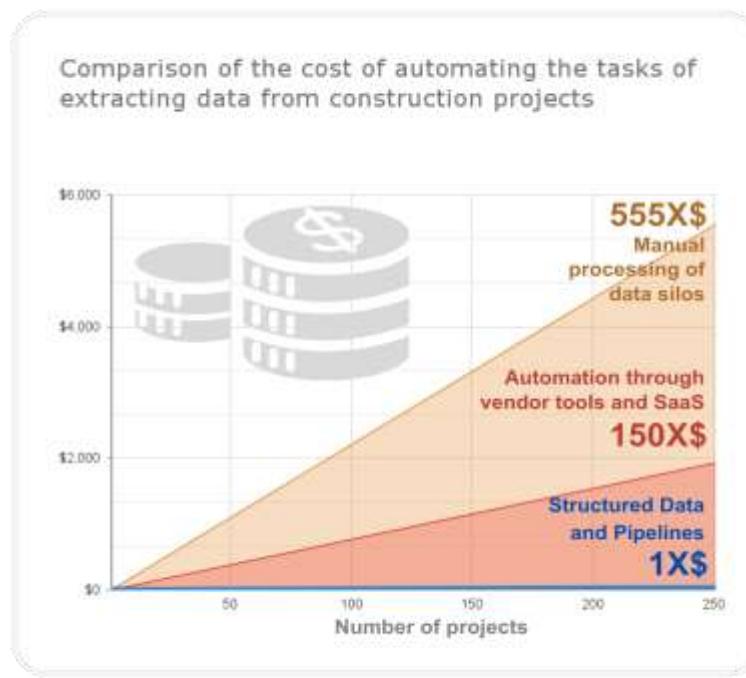
Данните от CAD системите и базите данни (BIM) са едни от най-усъвършенстваните и динамично актуализирани източници на данни в дейността на строителните компании. Тези приложения не само описват проекта с помощта на геометрия, но и го допълват с множество слоеве текстова информация: обеми, свойства на материалите, разпределение на помещенията, нива на енергийна ефективност, допуски, продължителност на живота и други атрибути.

Атрибутите, присвоени на обекти в CAD -моделите, се формират на етапа на проектиране и стават основа за по-нататъшни бизнес процеси, включително изчисляване на разходите, планиране, оценка на жизнения цикъл и интеграция с ERP и CAFM -системи, където ефективността на процесите до голяма степен зависи от качеството на данните, идващи от отделите за проектиране.

Традиционният подход за валидиране на атрибути в CAD- (BIM-) моделите включва ръчно валидиране (фиг. 7.2-1), което се превръща в дълъг и скъп процес, когато обемът на моделите е голям. Като се имат предвид обемът и броят на съвременните строителни проекти и техните редовни актуализации, процесът на валидиране и преобразуване на данните става неустойчив и недостъпен.

Главните изпълнители и ръководителите на проекти се сблъскват с необходимостта да обработват големи количества данни за проекти, включително множество версии и фрагменти на едни и същи модели. Данните идват от проектантските организации във формати RVT, DWG, DGN, IFC, NWD и други (фигура 3.1-14) и изискват редовна проверка за съответствие с индустриталните и корпоративните стандарти.

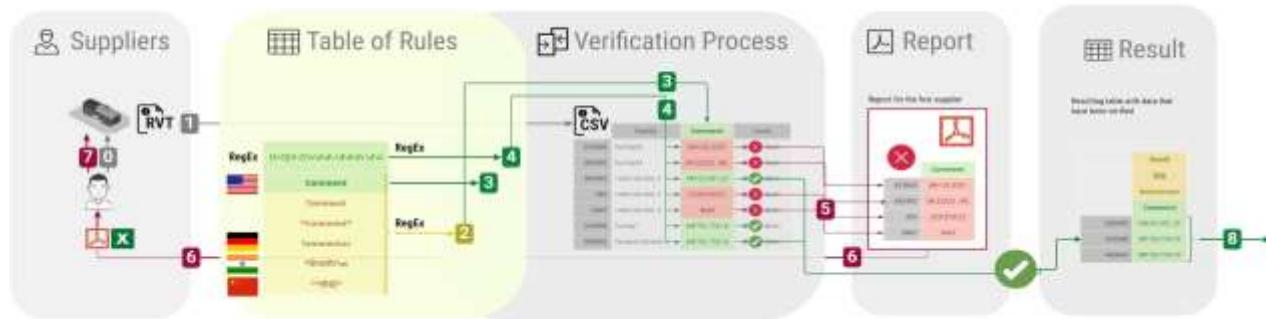
Зависимостта от ръчни действия и специализиран софтуер превръща процеса на валидиране на данни в пречка в работните процеси, свързани с данни от модели, използвани в цялата компания. Автоматизацията и използването на структурирани изисквания могат да елиминират тази зависимост, като увеличат значително скоростта и надеждността на валидирането на данните (фигура 7.3-7).



**Фиг. 7.3-7 Автоматизацията увеличава скоростта на проверка и обработка на данните, което намалява разходите за работа десетки пъти [140].**

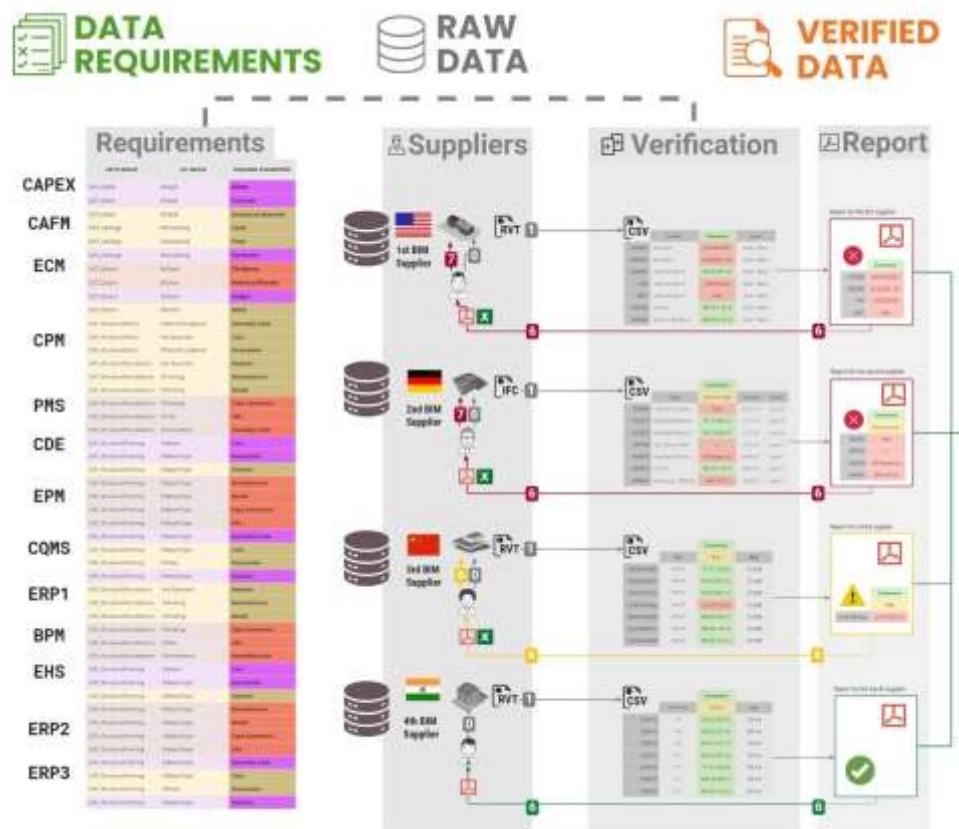
Процесът на валидиране на CAD данни включва извлечение на данни (ETL етап Extract) от различни затворени (RVT, DWG, DGN, NWD и т.н.) или отворени полуструктурни и параметрични формати (IFC, CPXML, USD.) или отворени полуструктурни и параметрични формати (IFC, CPXML, USD), в които могат да се прилагат таблици с правила към всеки атрибут и неговите стойности (етап Transform) с помощта на регулярни изрази RegEx (фиг. 7.3-8) - процес, който разглеждахме подробно в четвъртата част на книгата.

Създаването на PDF доклад за грешки на и успешно валидирани записи трябва да бъде завършено с изход (стъпка Load) в структурирани формати, които разглеждат само валидирани обекти, които могат да се използват за по-нататъшни процеси.



Фигура 7.3-8 Процесът на валидиране на данните от доставчиците на данни за проекта до окончателния доклад се валидира с помощта на регулярни изрази.

Автоматизирането на валидирането на данни от CAD системи (BIM) със структурирани изисквания и стрийминг на нови данни, които се обработват чрез ETL-Pipelines (фигура 7.3-9), намалява необходимостта от ръчно участие в процеса на валидиране (всеки от процесите на валидиране и изисквания към данните е разгледан в предишните глави).



Фиг. 7.3-9 Автоматизирането на валидирането на данни чрез ETL опростява управлението на строителни проекти, като ускорява процесите.

Традиционно валидирането на модели, предоставени от изпълнители и CAD (BIM) специалисти,

може да отнеме дни до седмици. С въвеждането на автоматизирани ETL процеси обаче това може да се сведе до няколко минути. В една типична ситуация изпълнителят заявява: "Моделът е валидиран и отговаря на изискванията." Това изявление поставя началото на верига от проверки на твърдението на изпълнителя за качеството на данните:

- ⌚ Ръководител на проекта - "Изпълнителят заявява: "Моделът е тестван, всичко е наред"."
- ⌚ Мениджър на данни - валидиране на зареждането:
  - Прост скрипт в Pandas открива нарушението за секунди. Автоматизацията елиминира споровете:
    - Категория: OST\_StructuralColumns, Параметър: FireRating IS NULL.
    - Генериране на списък с идентификатори на нарушения → експортиране в Excel/PDF.

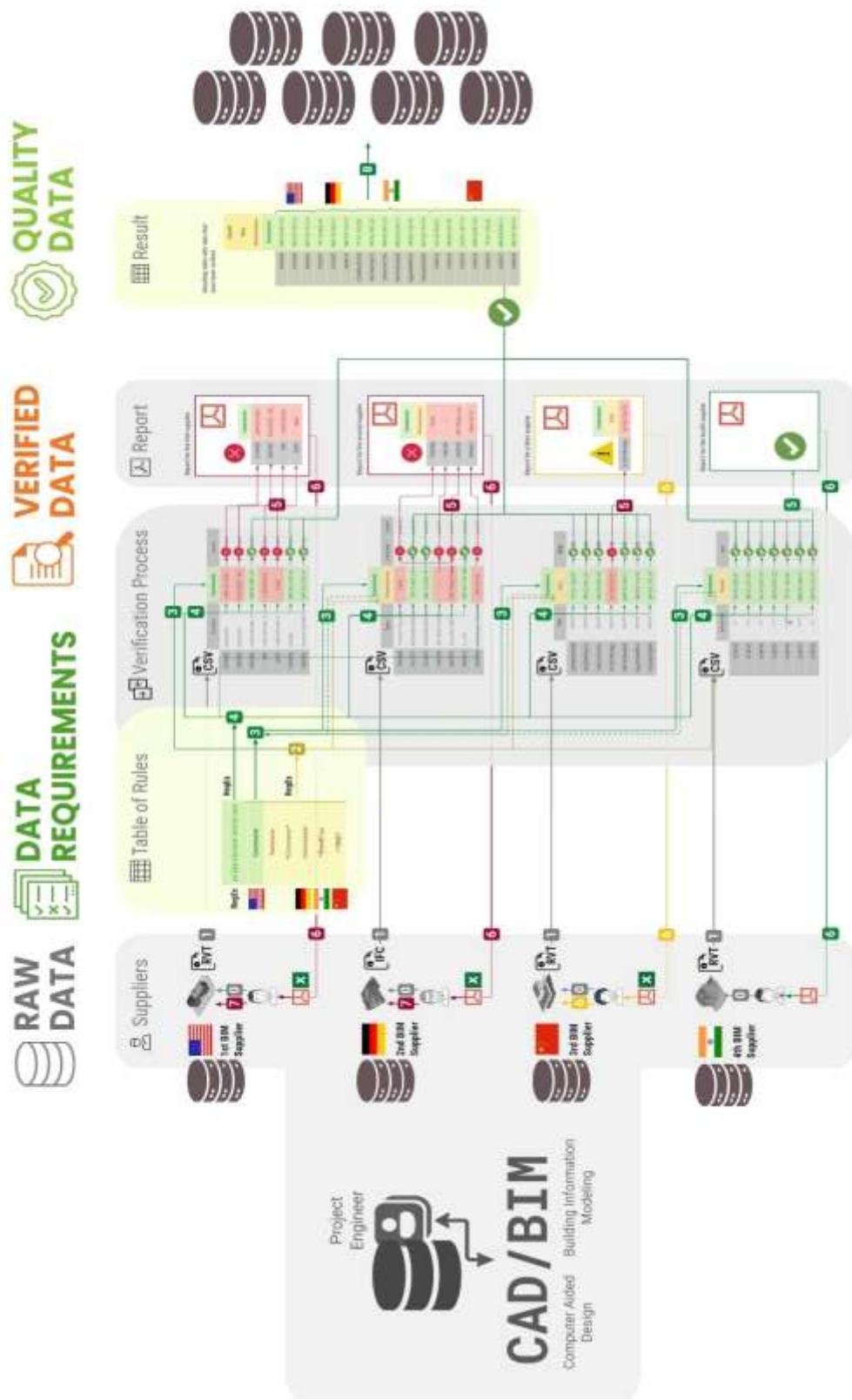
Един прост скрипт в Pandas открива нарушението за секунди:

```
df = model_data[model_data["Category"] == "OST_StructuralColumns"] # Филтриране
issues = df[df["FireRating"].isnull()] # Празни стойности
issues[["ElementID"]].to_excel("fire_rating_issues.xlsx") # Експорт на идентификатори
```

- ⌚ Мениджърът на данни към мениджъра на проекта - "Проверка на показва, че в 18 колони не е попълнен параметърът FireRating."
- ⌚ Ръководител на проекта към изпълнителя - "Моделът се връща за преработка: параметърът FireRating е задължителен, без него приемането е невъзможно"

В резултат на това CAD моделът не се подлага на валидиране, автоматизацията елиминира споровете, а изпълнителят почти незабавно получава структуриран доклад със списък на идентификаторите на проблемните елементи. По този начин процесът на валидиране става прозрачен, повторяем и защитен от човешки грешки (фиг. 7.3-10).

Този подход превръща процеса на валидиране на данните в инженерна функция, а не в ръчен контрол на качеството. Това не само повишава производителността, но и дава възможност за прилагане на една и съща логика към всички проекти на компанията, което позволява цялостна цифрова трансформация на процесите - от проектирането до операциите.



Фиг. 7.3-10 Автоматизирането на проверката на атрибутите на елементите елиминира човешката грешка и намалява вероятността от грешки.

Чрез използването на автоматизирани тръбопроводи (фиг. 7.3-10) потребителите на системата, които очакват качествени данни от системите CAD- (BIM-), могат незабавно да получат необходимите им изходни данни - таблици, документи, изображения - и бързо да ги интегрират в работните си задачи.

Автоматизацията на контрола, обработката и анализа води до промяна в подхода към управлението на строителните проекти, особено към оперативната съвместимост на различните системи, без да се използват сложни и скъпи модулни патентовани системи или затворени решения на доставчиците.

Докато понятията и маркетинговите съкращения идват и си отиват, самите процеси на валидиране на изискванията към данните завинаги ще останат неразделна част от бизнес процесите. Вместо да се създават все по-специализирани формати и стандарти, строителният бранш трябва да се обърне към инструменти, които вече са доказали своята ефективност в други отрасли. Днес съществуват мощни платформи за автоматизиране на обработката на данни и интегрирането на процеси, които позволяват на компаниите значително да намалят времето, отделяно за рутинни операции, и да сведат до минимум грешките при Extract, Transform и Load.

Един от популярните примери за решения за автоматизация и оркестрация на процесите на ETL е Apache Airflow, който позволява да се организират сложни изчислителни процеси и да се управляват конвейери на ETL. Наред с Airflow активно се използват и други подобни решения, като Apache NiFi за маршрутизиране и стрийминг на данни и n8n за автоматизация на бизнес процеси.



## ГЛАВА 7.4.

# ОРКЕСТРАЦИЯ НА ETL И РАБОТНИ ПОТОЦИ: ПРАКТИЧЕСКИ РЕШЕНИЯ

## DAG и Apache Airflow: автоматизация на работния процес и оркестрация

Apache Airflow е безплатна платформа с отворен код, предназначена за автоматизиране, оркестириране и наблюдение на работни потоци (ETL -конвейери).

Работата с големи количества данни се налага всеки ден:

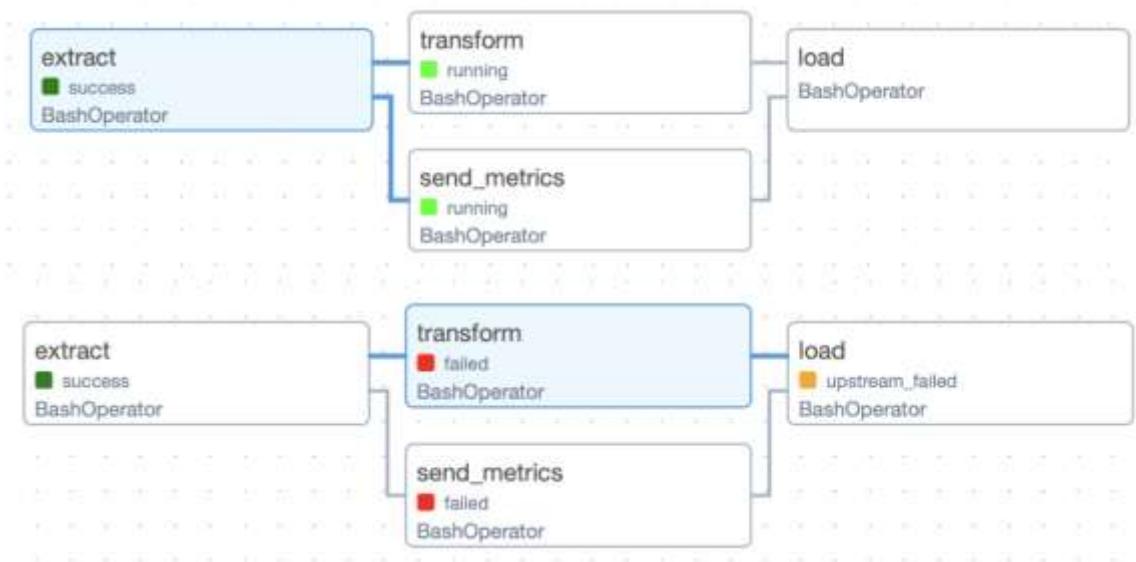
- Изтегляне на файлове от различни източници - Извличане на (например от доставчици или клиенти).
- Трансформиране на тези данни в необходимия формат - Трансформиране на (структурите, изчистване и валидиране)
- Изпращане на резултати за проверка и създаване на отчети - Зареждане на (качване в необходимите системи, документи, бази данни или информационни табла).

Ръчното изпълнение на такива ETL процеси отнема значително време и води до риск от човешка грешка. Промяна в източника на данни или неуспех на някоя от стъпките може да доведе до забавяне и неверни резултати.

Инструментите за автоматизация, като Apache Airflow, ви позволяват да изградите надежден ETL - конвейер, да сведете до минимум грешките, да намалите времето за обработка и да гарантирате, че данните са правилни на всеки етап. В основата на Apache Airflow е концепцията за DAG (Directed Acyclic Graph) - насочен ацикличен граф, в който всяка задача (оператор) е свързана с други зависимости и се изпълнява стриктно в определена последователност. DAG елиминира циклите, което осигурява логична и предвидима структура на изпълнението на задачите.

Airflow се грижи за оркестрацията - управление на зависимостите между задачите, контрол на графите за изпълнение, проследяване на състоянието и автоматично реагиране при неуспехи. Този подход свежда до минимум ръчната намеса и гарантира надеждността на целия процес.

**Оркестраторът на задачи** е инструмент или система, предназначена за управление и контрол на изпълнението на задачи в сложни компютърни и информационни среди. Той улеснява процеса на разгръщане, автоматизиране и управление на изпълнението на задачите с цел подобряване на производителността и оптимизиране на ресурсите.



**Фигура 7.4-1** Apache Airflow предоставя удобен за потребителя интерфейс, в който можете да визуализирате DAG -ETL, да преглеждате дневници за изпълнение, състояние на стартиране на задачите и др.

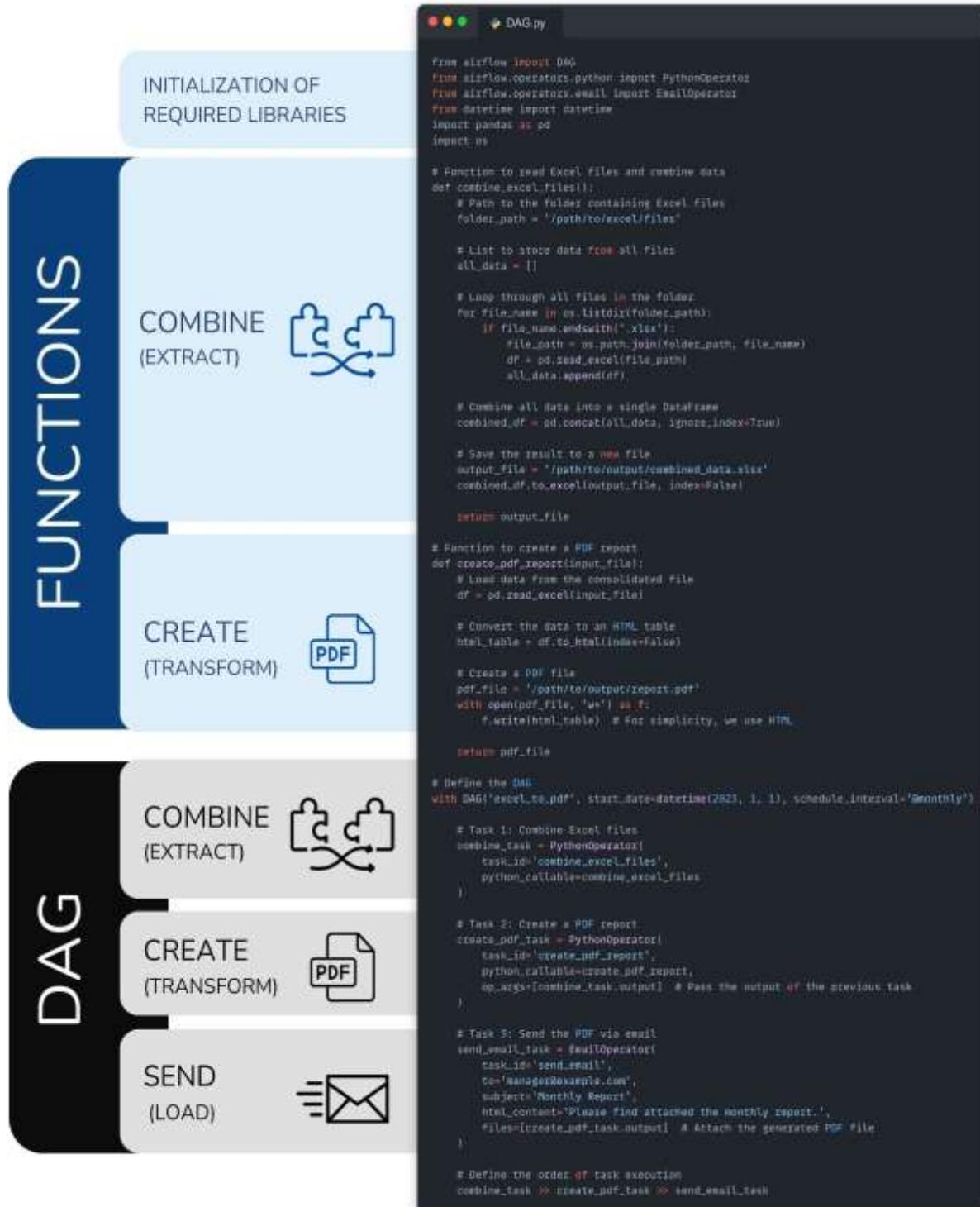
Airflow се използва широко за оркестириране и автоматизиране на разпределени изчисления, обработка на данни, управление на процеси ETL (Extract, Transform, Load), планиране на задачи и други сценарии за данни. По подразбиране Apache Airflow използва SQLite като база данни.

Пример за прост DAG, подобен на ETL, се състои от задачи - Извличане, Преобразуване и Зареждане. В графа, който се контролира чрез потребителски интерфейс (фиг. 7.4-1), се определя редът на изпълнение на задачите (фрагменти от кода): например първо се изпълнява extract (извличане), след това transform (и sending\_metrics), а задачата load (зареждане) завършва работата. Когато всички задачи са изпълнени, процесът на зареждане на данни се счита за успешен.

## Apache Airflow: практическо приложение автоматизация на ETL

Apache Airflow се използва широко за организиране на сложни процеси за обработка на данни, като позволява да се изграждат гъвкави ETL -конвейери. Apache Airflow може да се стартира или чрез уеб интерфейс, или програмно чрез код на Python (фиг. 7.4-2). В уеб интерфейса (фиг. 7.4-3) администраторите и разработчиците могат визуално да наблюдават DAG, да изпълняват задачи и да анализират резултатите от изпълнението.

С помощта на DAG, можете да зададете ясна последователност на задачите, да управлявате зависимости между тях и да реагирате автоматично на промените в изходните данни. Нека разгледаме пример за използване на Airflow за автоматизиране на обработката на отчети (фиг. 7.4-2).



Фиг. 7.4-2 Концепция за ETL -конвейер за обработка на данни с помощта на Apache Airflow.

ETL (EXTRACT, TRANSFORM, LOAD) PIPELINE

В този пример (Фигура 7.4-2) се разглежда DAG, която изпълнява ключови задачи в рамките на ETL -конвейера:

**■ Прочетете Excel -файлове (Извлечете):**

- Последователно обхождане на всички файлове в дадена директория.
- Прочитане на данни от всеки файл с помощта на библиотеката pandas.
- Обединяване на всички данни в една DataFrame.

**■ Създаване на PDF -документ (Transform):**

- Преобразува обединената DataFrame в HTML -таблица.
- Запишете таблицата като PDF (в демо версията - чрез HTML).

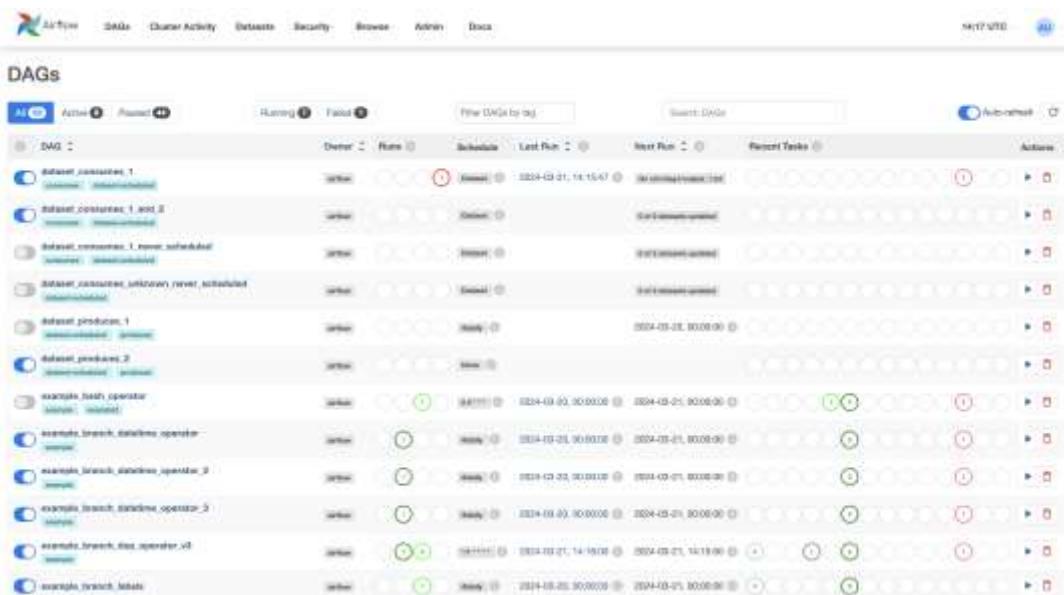
**■ Изпращане на отчет по електронна поща (Load):**

- Приложете EmailOperator, за да изпратите PDF -документ по имейл.

**■ Конфигуриране на DAG:**

- Определяне на последователността на задачите: извлечение на данни → генериране на отчет → изпращане.
- Присвояване на график за стартиране (@monthly (ежемесечно) - първият ден от всеки месец).

Автоматизираният пример за ETL (Фигура 7.4-2) показва как да се съберат данни от Excel -файлове, да се създаде PDF -документ и да се изпрати по имейл. Това е само един от многото възможни случаи на използване на Airflow. Този пример може да бъде адаптиран към всяка конкретна задача, за да се опрости и автоматизира обработката на данни.



Фиг. 7.4-3 Преглед на всички DAGs в средата с информация за последните изпълнения.

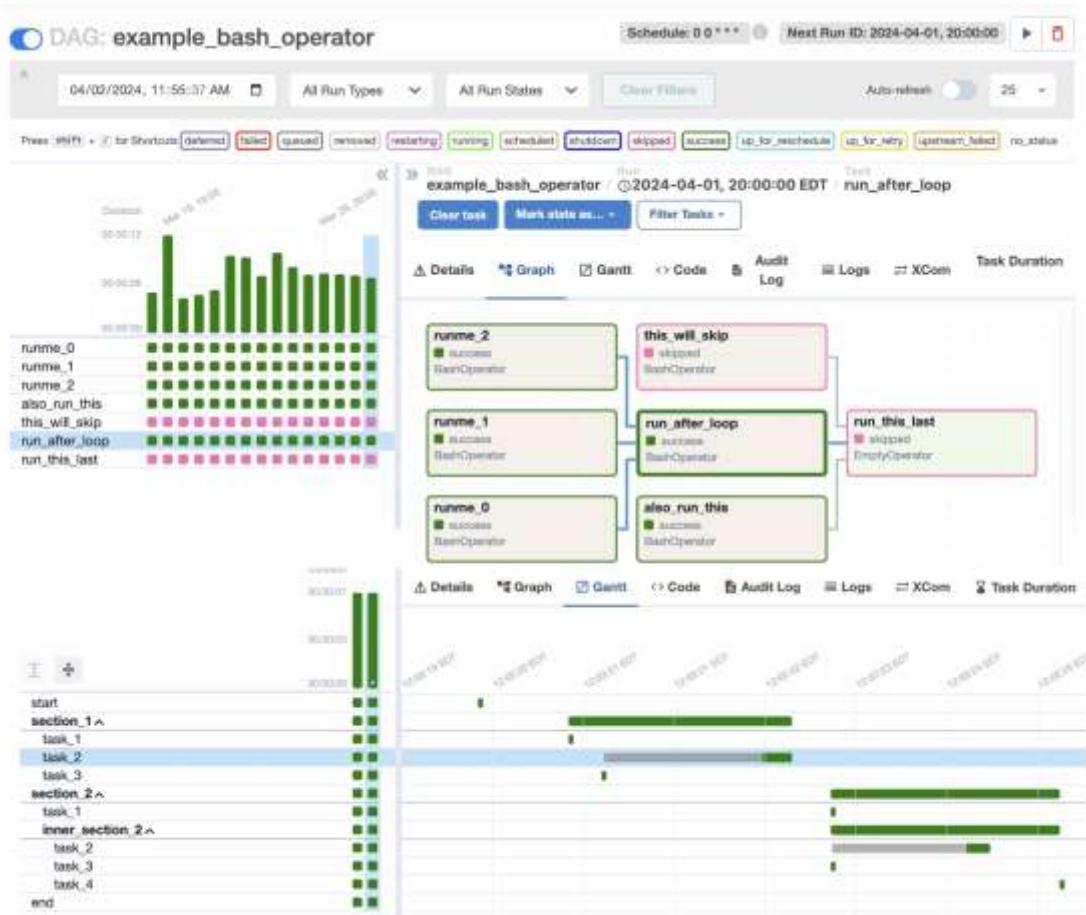
Уеб интерфейсът на Apache Airflow (Фигура 7.4-3) предоставя цялостна визуална среда за управление на работните потоци от данни. Той показва DAG като интерактивни графики, в които възлите представляват задачи, а ребрата - зависимости между тях, което улеснява проследяването на сложни работни потоци от данни. Интерфейсът включва табло за управление с информация за

състоянието на изпълнение на задачите, история на изпълнението, подробни регистри и показатели за производителност. Администраторите могат ръчно да стартират задачи, да рестартират неуспешни операции, да спират DAGs и да конфигурират променливи на средата - всичко това чрез интуитивен потребителски интерфейс.

Такава архитектура може да бъде допълнена с валидиране на данни, известия за състоянието на изпълнение, интеграция с външни API или бази данни. Airflow позволява гъвкаво адаптиране на DAG: добавяне на нови задачи, промяна на реда им, комбиниране на вериги - което го прави ефективен инструмент за автоматизиране на сложни процеси за обработка на данни. Когато стартирате DAG в уеб интерфейса на Airflow (фиг. 7.4-3, фиг. 7.4-4), можете да следите състоянието на задачите. Системата използва цветна индикация:

- Зелено - задачата е изпълнена успешно.
  - Жълто - процесът е в ход.
  - Червено - грешка при изпълнението на задачата.

В случай на неизправност (напр. липсващ файл или нарушена структура на данните) системата автоматично започва да изпраща известие.



Фигура 7.4-4 Apache Airflow значително опростява диагностиката на проблеми, оптимизацията на процеси и сътрудничеството в екип при сложни конвейери за обработка на данни.

Apache Airflow е удобен, тъй като автоматизира рутинните задачи, като премахва необходимостта да ги изпълнявате ръчно. Той гарантира надеждност, като следи изпълнението на процесите и осигурява незабавно уведомяване за грешки. Гъвкавостта на системата позволява лесно да се добавят нови задачи или да се променят съществуващите, като работните процеси се адаптират към променящите се изисквания.

В допълнение към Apache Airflow, съществуват подобни инструменти за оркестриране на работни потоци. Например безплатният Prefect с отворен код (фигура 7.3-5) предлага по-опростен синтаксис и се интегрира по-добре с Python, Luigi, разработен от Spotify, осигурява подобна функционалност и работи добре с големи данни. Заслужава да се отбележат и Kronos и Dagster, които предлагат съвременни подходи за изграждане на Pipeline с акцент върху модулността и мащабируемостта. Изборът на инструмент за оркестриране на задачи зависи от конкретните нужди на проекта, но всички те помагат за автоматизиране на сложни ETL процеси за данни.

Особено внимание заслужава Apache NiFi, платформа с отворен код, предназначена за поточно предаване и маршрутизиране на данни. За разлика от Airflow, която се фокусира върху пакетна обработка и управление на зависимостите, NiFi се фокусира върху преобразуването на данни в реално време, в движение и гъвкаво маршрутизиране между системите.

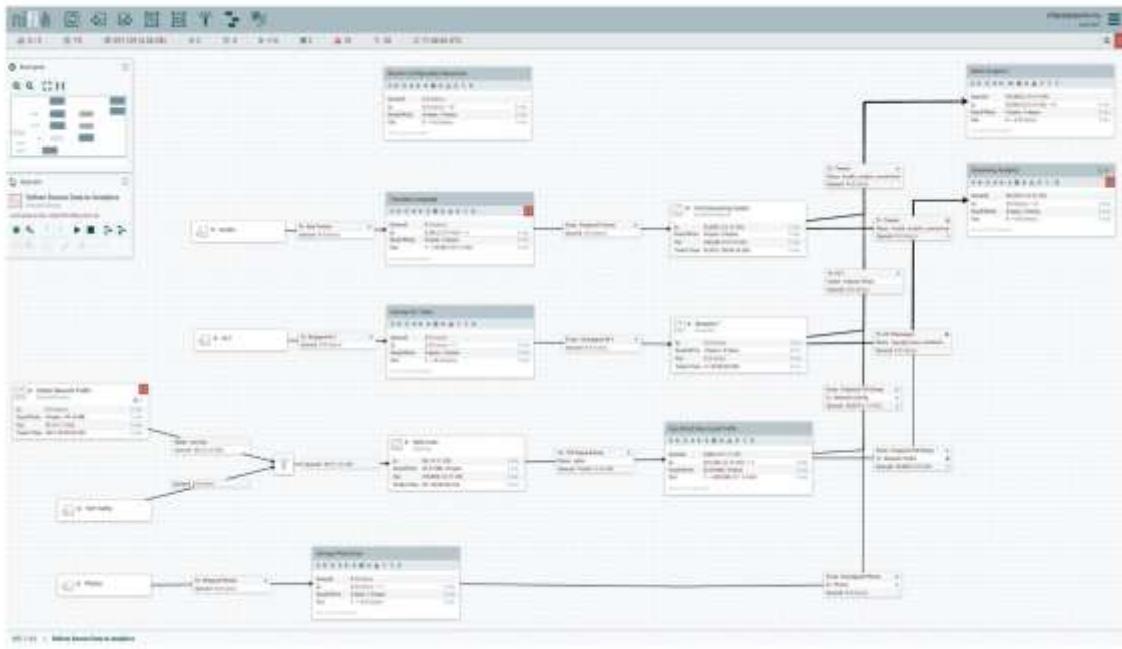
## Apache NiFi за маршрутизиране и преобразуване на данни

Apache NiFi е мощна платформа с отворен код, предназначена за автоматизиране на потоците от данни между различни системи. Първоначално е разработена през 2006 г. от Агенцията за национална сигурност на САЩ (NSA) под името "Niagara Files" за вътрешна употреба. През 2014 г. проектът е с отворен код и е прехвърлен на Apache Software Foundation, като става част от техните инициативи за трансфер на технологии [141].

Apache NiFi е проектиран да събира, обработва и предава данни в реално време. За разлика от Airflow, който работи с пакетни задачи и изисква добре дефинирани графики, NiFi работи в режим на поточно обработване, което позволява непрекъснато прехвърляне на данни между различни услуги.

Apache NiFi е идеален за интеграция с IoT устройства, строителни сензори, системи за наблюдение и например за поточно валидиране на CAD формати на сървър, където може да се изисква незабавна реакция на промени в данните.

С вградените инструменти за филтриране, трансформиране и маршрутизиране NiFi позволява данните да бъдат стандартизиирани (Transform), преди да бъдат прехвърлени (Load) към системите за съхранение или анализ. Едно от основните му предимства е вградената поддръжка за сигурност и контрол на достъпа, което го прави надеждно решение за обработка на чувствителна информация.



Фигура 7.4-5 Графично представяне на потока от данни в интерфейса Apache NiFi.

Apache NiFi ефективно се справя със задачите за поточно предаване, филтриране и маршрутизиране на данни в реално време. Той е идеален за технически интензивни сценарии, при които са важни стабилният трансфер на информация между системите и високата пропускателна способност.

Въпреки това, когато основната цел е да се интегрират различни услуги, да се автоматизират рутинни операции и бързо да се създадат работни потоци без задълбочени познания по програмиране, се търсят решения с нисък праг на достъп и максимална гъвкавост. Един такъв инструмент е n8n - платформа от клас Low-Code /No-Code, фокусирана върху автоматизацията на бизнеса и визуалната оркестрация на процесите.

n8n Low-Code, No-Code оркестрация на процеси

**n8n** е платформа с отворен код с нисък код / без код за изграждане на автоматизирани работни процеси, която се характеризира с лекота на използване, гъвкавост и възможност за бърза интеграция с широк набор от външни услуги.

**No-Code** е метод за създаване на цифрови продукти без писане на код. Всички елементи на процеса - от логиката до интерфейса - се реализират изключително с помощта на визуални инструменти. No-Code платформите са насочени към потребители без технически опит и позволяват бързо създаване на автоматизации, формуляри, интеграции и уеб приложения. Пример: потребителят създава автоматично изпращане на известия или интеграция с Google Sheets чрез интерфейс drag-and-drop, без да има познания по програмиране.

С възможностите за отворен код и локално внедряване n8n в процесите на автоматизация и създаване на ETL тръбопроводи дава на компаниите пълен контрол върху техните данни, като същевременно гарантира сигурност и независимост от доставчиците на облачни услуги.

За разлика от Apache Airflow, който е ориентиран към изчислителни задачи със строга оркестрация и изиска познания по Python, n8n предоставя визуален редактор, който позволява писане на скриптове, без да се изискват познания по езици за програмиране (фигура 7.4-6). Въпреки че интерфейсът му позволява създаването на автоматизирани процеси без писане на код (No-Code), при по-сложни сценарии потребителите могат да добавят свои собствени JavaScript и Python -функции, за да разширят възможностите (Low-Code).

**Low-Code** е подход за разработване на софтуер, при който основната логика на дадено приложение или процес се създава с помощта на графичен интерфейс и визуални елементи, а програмният код се използва само за персонализиране или разширяване на функционалността. Low-Code платформите позволяват значително да се ускори разработването на решения, като се включват не само програмисти, но и бизнес потребители с основни технически умения. Пример: потребителят може да изгради бизнес процес от готови блокове и, ако е необходимо, да добави свой собствен скрипт на JavaScript или Python.

Въпреки че n8n се позиционира като платформа с нисък праг за влизане, са необходими основни познания по програмиране, разбиране на уеб технологиите и умения за работа с API. Гъвкавостта на системата позволява тя да бъде адаптирана към широк спектър от задачи - от автоматизирана обработка на данни до интеграция с месинджъри, IoT -устройства и облачни услуги.

Основни характеристики и ползи от използването на n8n:

- Опциите за внедряване с **отворен код** и локално внедряване осигуряват пълен контрол на данните, съответствие със сигурността и независимост от доставчиците на облачни услуги.
- **Интеграция с над 330 услуги**, включително CRM, ERP, електронна търговия, облачни платформи, месинджъри и бази данни.
- **Гъвкавост на сценарии**: от прости известия до сложни вериги с API обработка на -заявки, логика за вземане на решения и свързване на AI -услуги.
- **Поддръжка на JavaScript и Python**: потребителите могат да вграждат потребителски код според нуждите си, като разширяват възможностите за автоматизация.
- **Интуитивен визуален интерфейс**: позволява бързо да конфигурирате и визуализирате всички стъпки на процеса.

Платформи от клас Low-Code предоставят инструменти за създаване на цифрови решения с минимален код, което ги прави идеални за екипи, които нямат задълбочени технически познания, но трябва да автоматизират процеси.

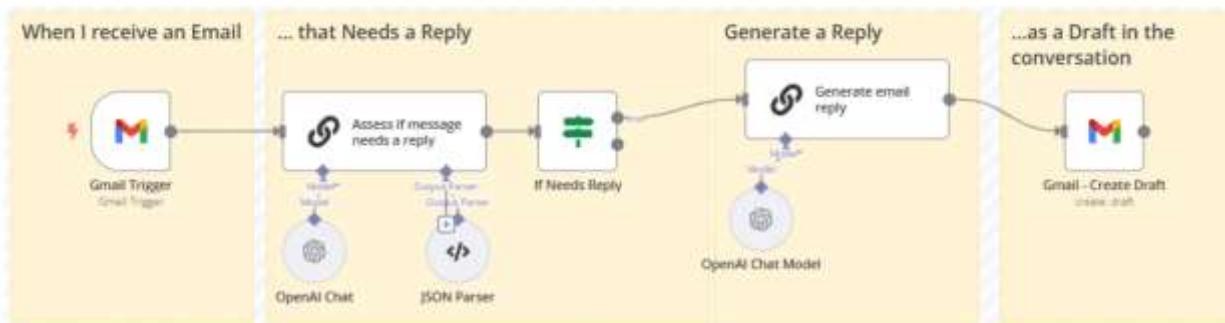
В строителството n8n може да се използва за автоматизиране на различни процеси, като например интегриране със системи за управление на проекти, проверка на потоци, писане на готови отчети и писма, автоматично актуализиране на данните за материалните запаси, изпращане на уведомления за състоянието на задачите до екипите и др. Персонализираният тръбопровод в n8n може да

намали многократно ръчните операции, да намали вероятността от грешки и да ускори вземането на решения за изпълнение на проекта.

Можете да избирате от почти две хиляди готови, безплатни и с отворен код n8n Pipeline, достъпни на адрес: [n8n.io/workflows](https://n8n.io/workflows), за да автоматизирате както строителните работни процеси, така и личните задачи, като намалите рутинните операции.

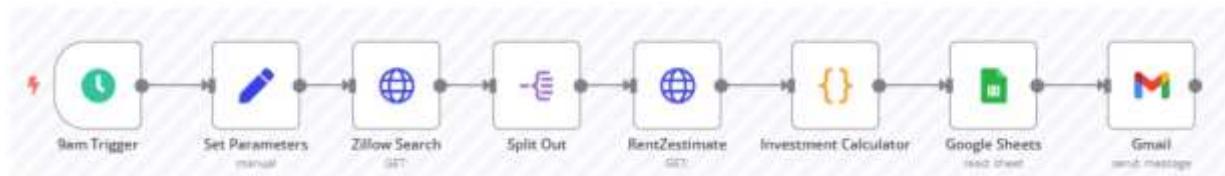
Вземете един от готовите шаблони на Pipeline, достъпни бесплатно в [n8n.io](https://n8n.io) [142], който автоматично създава чернови на отговори в Gmail (фиг. 7.4-6), като помага на потребителите, които получават голям обем имейли или изпитват затруднения при съставянето на отговори.

Този шаблон n8n "Gmail AI Auto-Responder: Create draft responses to incoming emails" (Фигура 7.4-6) анализира входящите съобщения, като използва LLM от ChatGPT, определя дали е необходим отговор, генерира чернова от ChatGPT и преобразува текста в HTML и го добавя към веригата от съобщения в Gmail. При това имейлът не се изпраща автоматично, което ви позволява ръчно да редактирате и одобрите отговора. Настройката отнема около 10 минути и включва OAuth конфигурация на Gmail API и OpenAI API интеграция. Резултатът е удобно и бесплатно решение за автоматизиране на рутинната имайл комуникация, без да се губи контролът върху съдържанието на имейлите.



Фиг. 7.4-6 Автоматизиран процес на генериране на отговор на електронна поща с помощта на n8n.

Друг пример за автоматизация с n8n е намирането на изгодни сделки на пазара на недвижими имоти [143]. N8n Pipeline "Automate daily property deals using Zillow API, Google Sheets and Gmail", събира ежедневно подходящи оферти, отговарящи на зададени критерии, като използва Zillow API. Той автоматично изчислява ключови инвестиционни показатели (възвръщаемост на инвестициите в брой, месечен паричен поток, първоначална вноска), актуализира Google Sheets и изпраща обобщен доклад по електронна поща (фигура 7.4-7), което позволява на инвеститорите да спестят време и да отговорят бързо на най-добрите оферти.



Фиг. 7.4-7 Автоматизиран процес за оценка на инвестиционната привлекателност на недвижими имоти.

Неговата гъвкавост и разширяемост правят n8n ценен инструмент за компаниите, които се стремят да се трансформират в цифрова среда и да станат по-конкурентоспособни на пазара със сравнително прости и безплатни инструменти с отворен код.

Инструменти като Apache NiFi, Airflow и n8n могат да се разглеждат като три слоя за обработка на данни (фигура 7.4-8). NiFi управлява потока от данни, като осигурява тяхното доставяне и преобразуване, Airflow организира изпълнението на задачите, като обединява данните в конвейери за обработка, а n8n автоматизира интеграцията с външни услуги и управлява бизнес логиката.



	The main task	Approach
Apache NiFi	Streaming and data transformation	Real-time stream processing
Apache Airflow	Task orchestration, ETL pipelines	Batch planning, DAG processes
n8n	Integration, automation of business logic	Low-code visual orchestration

Фиг. 7.4-8 Apache Airflow, Apache NiFi и n8n могат да се разглеждат като три допълващи се слоя на съвременната архитектура за управление на данни.

Заедно тези безплатни инструменти с отворен код могат да бъдат пример за ефективна екосистема за управление на данни и процеси в строителната индустрия, която позволява на компаниите да използват информацията за вземане на решения и автоматизиране на процесите.

## Следващи стъпки: преминаване от ръчни операции към аналитични решения

Съвременните строителни компании работят в среда на висока несигурност: променящи се цени на материалите, забавени доставки, недостиг на работна ръка и кратки срокове за изпълнение на проектите. Използването на аналитични табла, ETL -конвейери и BI системи помага на компаниите бързо да идентифицират проблемните области, да оценяват ефективността на ресурсите и да предвиждат промените преди те да доведат до финансови загуби.

За да обобщим тази част, си струва да подчертаем основните практически стъпки, които ще ви помогнат да прилагате обсъжданите технологии в ежедневните си задачи:

- Внедряване на визуализации на данни и аналитични табла
  - Усвояване на процеса на създаване на информационни табла за наблюдение на ключови показатели за ефективност (KPI)
  - Използвайте инструменти за визуализация на данните (Power BI, Tableau, Matplotlib,

Plotly).

■ Автоматизиране на обработката на данни чрез ETL -процеси

- Създаване на автоматично събиране на данни от различни източници (документация, таблици, CAD) чрез ETL процеси
- Организиране на преобразуване на данни (напр. проверка на регулярни изрази или изчисления) с помощта на скриптове на Python
- Опитайте се да настроите автоматични PDF (или DOC) отчети с библиотеката FPDF, като използвате данни от файлове на Excel или извличате информация от други PDF документи.

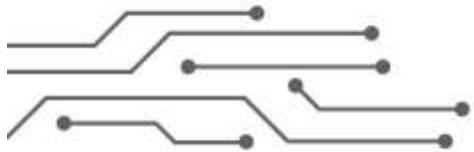
■ Използване на езикови модели (LLM) за автоматизация

- Използване на големи езикови модели (LLM) за генериране на код, който помага за извличане и анализиране на данни от неструктурирани документи.
- Запознайте се с инструмента за автоматизация n8n и разгледайте готовите шаблони и казуси на техния уебсайт. Определете кои процеси от вашата работа могат да бъдат напълно автоматизирани, като използвате подхода No-Code/Low-Code

Аналитичният подход към данните и автоматизацията на процесите не само намалява времето, отделяно за рутинни операции, но и подобрява качеството на вземане на решения. Компаниите, които внедряват инструменти за визуален анализ и ETL -конвейери, получават възможност да реагират бързо на промените

Автоматизирането на бизнес процесите с помощта на инструменти като n8n, Airflow и NiFi е само първата стъпка към цифровата зрялост. Следващата стъпка е качественото съхранение и управление на самите данни, които са в основата на автоматизацията. В част 8 разглеждаме подробно как строителните компании могат да изградят устойчива архитектура за съхранение на данни, като преминат от хаос от документи и многоформатни файлове към централизирани платформи за съхранение и анализ.





## VIII ЧАСТ

# СЪХРАНЕНИЕ И УПРАВЛЕНИЕ НА ДАННИ В СТРОИТЕЛСТВОТО

Част 8 разглежда съвременните технологии за съхранение и управление на данни в строителната индустрия. В нея се анализират ефективни формати за обработка на големи обеми информация - от обикновените CSV и XLSX до по-производителните Apache Parquet и ORC с подробно сравнение на техните възможности и ограничения. Разглеждат се концепциите за складове за данни (DWH), езера от данни) и техните хибридни решения (Data Lakehouse), както и принципите за управление на данните) и минимализъм на данните). Подробно са разгледани проблемите на блатото от данни) и стратегиите за предотвратяване на хаоса в информационните системи. Представени са нови подходи за работа с данни, включително векторни бази данни и тяхното приложение в строителството чрез концепцията Bounding Box. В тази част са засегнати и методологията DataOps и VectorOps като нови стандарти за организиране на работните процеси с данни.

## ГЛАВА 8.1.

# ИНФРАСТРУКТУРА ЗА ДАННИ: ОТ ФОРМАТИ ЗА СЪХРАНЕНИЕ ДО ЦИФРОВИ ХРАНИЛИЩА

### Атомите на данни: основата на ефективното управление на информацията

Всичко във Вселената се състои от най-малките градивни елементи - атоми и молекули, и с течение на времето всички живи и неживи същества неизбежно се връщат към това първоначално състояние. В природата този процес протича с изумителна скорост, която ние се опитваме да пренесем върху процесите, контролирани от човека.

В гората всички живи организми в крайна сметка се превръщат в хранително вещество, което служи за основа на нови растения. Тези растения на свой ред се превръщат в храна за нови живи същества, съставени от същите атоми, които са създали Вселената преди милиони години.

В света на бизнеса също е важно да се разбиват сложни, многопластови структури на техните най-основни, минимално обработени единици - подобно на атомите и молекулите в природата. Това позволява атомите от данни да бъдат ефективно съхранявани и управлявани, превръщайки ги в богата, плодородна основа, която се превръща в ключов ресурс за растежа на анализа и качеството на решенията.



Фиг. 8.1-1 Анализът и вземането на решения се основават на повторно използвани данни, които веднъж са били обработени и съхранени.

Музикалните композиции се състоят от ноти, които се комбинират, за да създадат сложни музикални произведения, докато думите се създават от една примитивна единица - буквата и звука. Независимо дали става въпрос за природата, науката, икономиката, изкуството или технологиите, светът показва забележително единство и хармония в стремежа си към разрушение, структура, цикличност и съзидание. По подобен начин процесите в системите за калкулиране на разходите се разбиват на малки структурирани единици - ресурсни елементи - на ниво калкулиране на разходите и графици. Тези единици, подобно на бележките, след това се използват за формиране на по-сложни калкулации и графици. Същият принцип се използва от системите за автоматизирано проектиране, в които сложните архитектурни и инженерни проекти се изграждат от основни елементи - отделни елементи и библиотечни компоненти, от които се създава пълен 3D-модел на проекта на сложна сграда или структура.

Концепцията за цикличност и структура, присъща на природата и науката, се отразява и в съвременния свят на данните. Точно както в природата всички живи същества се връщат към атомите и молекулите, така и в света на съвременните инструменти за обработка на данни информацията има тенденция да се връща към най-примитивната си форма.

Най-малките елементи с тяхната крайна неделимост са основните градивни елементи на бизнес процесите. Важно е от самото начало внимателно да се обмисли как да се събират, структурират (разбиват на атоми) и съхраняват тези малки градивни елементи от различни източници. Организирането и съхраняването на данни не е просто въпрос на разбиването им на съставни части. Също толкова важно е да се гарантира, че те са интегрирани и съхранявани по структуриран начин, така че данните да могат лесно да бъдат извлечани, анализирани и използвани за вземане на решения, когато са необходими.

За да се обработва ефективно информацията, трябва внимателно да се подберат форматът и методите за съхранение на данни - точно както почвата трябва да се подготви за растежа на дърветата. Съхранението на данни трябва да бъде организирано така, че да се гарантира високо качество и релевантност на информацията, като се елиминират излишните или нерелевантните данни. Колкото по-добре е структурирана тази "информационна почва", толкова по-бързо и по-точно потребителите ще могат да намират правилните данни и да решават аналитични проблеми.

## Съхранение на информация: файлове или данни

Складовете за данни позволяват на компаниите да събират и комбинират информация от различни системи, създавайки единен център за по-нататъшен анализ. Събранныте исторически данни дават възможност не само за по-задълбочен анализ на процесите, но и за идентифициране на модели, които могат да повлият на бизнес резултатите.

Да речем, че дадена компания работи по няколко проекта едновременно. Инженерът иска да разбере колко бетон е излят и колко все още трябва да се закупи. При традиционния подход той би трявало ръчно да търси в сървъра и да отвори няколко таблици с оценки, да ги сравни с удостоверенията за изпълнени работи и да провери текущите наличности. Това отнема часове или дори дни. Дори и с ETL процесите и автоматичните скриптове задачата остава полуръчна: инженерът

все още трябва ръчно да посочва пътя до папките или конкретни файлове на сървъра. Това намалява цялостния ефект от автоматизацията, тъй като тя продължава да отнема ценно работно време.

При преминаване към управление на данни, вместо да работи с файловата система на сървъра, инженерът получава достъп до единна структура за съхранение, където информацията се актуализира в реално време. Една-единствена заявка - под формата на код, SQL -запитване или дори LLM - агентско повикване - може незабавно да предостави точни данни за текущите баланси, обемите на извършената работа и предстоящите доставки, ако данните са предварително подгответи и обединени в склад за данни, където не е необходимо да се рови из папки, да се отварят десетки файлове и ръчно да се сравняват стойности.

Дълго време строителните компании използваха PDF -документи, DWG -чертежи, RVT -модели и стотици и хиляди Excel -таблици и други разнородни формати, които се съхраняват в определени папки на сървърите на компанията, което затруднява търсенето на информация, нейната проверка и анализ. В резултат на това файловете, които остават след приключването на проектите, най-често се преместват обратно на сървъра в архивни папки, които на практика не се използват в бъдеще. Такова традиционно съхранение на данни, базирано на файлове, губи значение с увеличаването на потока от данни поради уязвимостта си към човешки грешки.

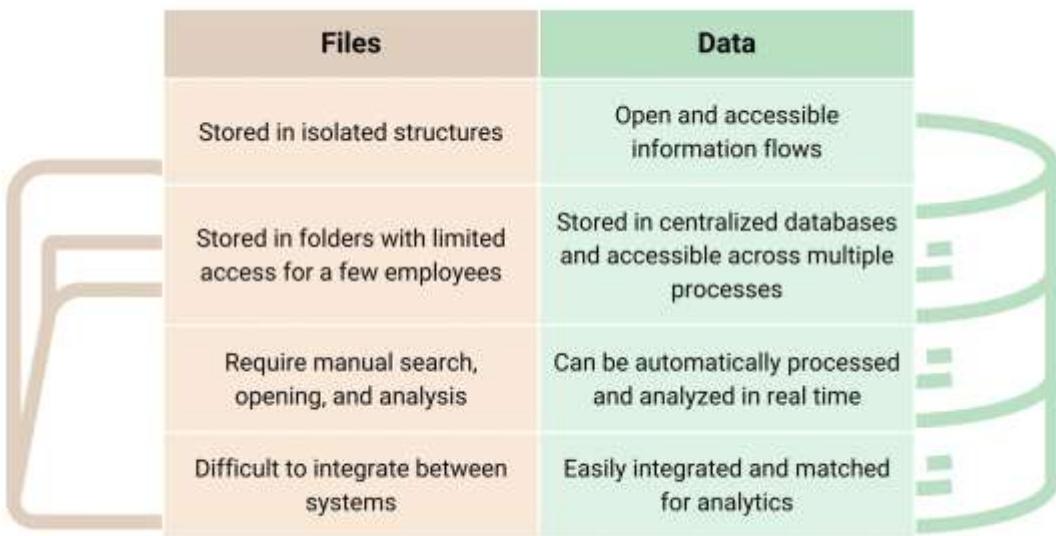
**Файлът** е просто изолиран контейнер, в който се съхраняват данни. Файловете се създават за хора, а не за системи, така че изискват ръчно отваряне, четене и интерпретиране. Примери за това са Excel -таблица, PDF -документ или CAD -чертеж, който трябва да бъде специално отворен в определен инструмент, за да се получи достъп до желаната информация. Без структурирано извлечане и обработка информацията в него остава неизползвана.

**Данните от** своя страна са машинно четима информация, която се свързва, актуализира и анализира автоматично. В един склад за данни (например база данни, DWH или езеро от данни) информацията е представена под формата на таблици, записи и връзки. Това дава възможност за унифицирано съхранение, автоматизирани заявки, стойностни анализи и отчитане в реално време.

Използването на данни вместо файлове (фиг. 8.1-1) елиминира процеса на ръчно търсене и унифицира обработката. Компаниите, които вече са приложили такъв подход, получават конкурентно предимство поради бързината на достъпа до информация и възможността за бързото ѝ интегриране в бизнес процесите.

Преминаването от използване на файлове към данни е неизбежна промяна, която ще определи бъдещето на строителната индустрия.

Всяка компания в строителния бранш ще бъде изправена пред ключов избор: да продължи да съхранява информация в разнородни файлове и силози, които трябва да се разчитат от хора с помощта на специални програми, или да я трансформира в първите етапи на обработка в структурирани данни, създавайки единна интегрирана цифрова основа за автоматизирано управление на проекти.



**Фигура 8.1-1 Еволюция на информационния поток: от изолирани файлове към интегрирани данни.**

С нарастващето на обема на информацията традиционните методи за съхранение и обработка на файлове стават все по-малко ефективни. В строителната индустрия, както и в други сектори, вече не е достатъчно да се разчита на разнородни папки с различни файлови формати или несвързани бази данни.

Фирмите, които искат да останат конкурентоспособни в цифровата ера, неизбежно ще се насочат към интегрирани цифрови платформи, ще използват технологии за големи обеми от данни и автоматизирани системи за анализ.

Преминаването от съхранение на файлове към работни процеси, базирани на данни, ще изисква преосмисляне на подходите за управление на информацията и съзнателен избор на формати, подходящи за по-нататъшно интегриране в централизирани хранилища. Този избор ще определи колко ефективно могат да се обработват данните, колко бързо може да се получи достъп до тях и колко лесно могат да се интегрират в цифровите процеси на компанията.

## Съхраняване на големи обеми данни: анализ на популярните формати и тяхната ефективност

Форматите за съхранение играят ключова роля за мащабируемостта, надеждността и производителността на инфраструктурата за анализ. За анализ и обработка на данни - като филтриране, групиране и агрегиране - в нашите примери се използва Pandas DataFrame - популярна структура за работа с данни в оперативната памет.

Въпреки това Pandas DataFrame няма собствен формат за съхранение, така че след приключване на обработката данните се експортират в един от външните формати - най-често CSV или XLSX. Тези таблични формати са лесни за обмен и съвместими с повечето външни системи, но имат редица ограничения: ниска ефективност на съхранение, липса на компресия и слаба поддръжка на

версии:

- **CSV** (стойности, разделени със запетая): прост текстов формат, който се поддържа от различни платформи и инструменти. Той е лесен за използване, но не поддържа сложни типове данни и компресия.
- **XLSX** (Excel Open XML Spreadsheet): файлов формат на Microsoft Excel, който поддържа сложни функции, като формули, диаграми и стилизиране. Въпреки че е полезен за ръчен анализ на данни и визуализация, той не е оптимизиран за мащабна обработка на данни.

В допълнение към популярните таблични формати XLSX и CSV, съществуват няколко популярни формата за ефективно съхранение на структурирани данни (фигура 8.1-2), като всеки от тях има уникални предимства в зависимост от специфичните изисквания за съхранение и анализ на данни:

- **Apache Parquet**: файлов формат за съхранение на данни в колони, оптимизиран за използване в системи за анализ на данни. Той предлага ефективни схеми за компресиране и кодиране на данни, което го прави идеален за сложни структури от данни и обработка на големи обеми от данни.
- **Apache ORC (Optimised Row Columnar)**: Подобно на Parquet, ORC осигурява висока степен на компресия и ефективно съхранение на данни. Той е оптимизиран за тежки операции за четене и е подходящ за съхранение на езера от данни.
- **JSON (JavaScript Object Notation)**: въпреки че JSON не е толкова ефективен по отношение на съхранението на данни в сравнение с двоични формати като Parquet или ORC, той е много достъпен и лесен за работа, което го прави идеален за скриптове, при които са важни четимостта и уеб съвместимостта.
- **Feather**: бърз, лек и лесен за използване аналитично ориентиран двоичен формат за съхранение на данни в колони. Той е проектиран за ефективно прехвърляне на данни между Python (Pandas) и R, което го прави отличен избор за проекти, включващи тези програмни среди.
- **HDF5 (Hierarchical Data Format version 5)**: предназначен за съхранение и организиране на големи количества данни. Той поддържа широк набор от типове данни и е подходящ за работа със сложни колекции от данни. HDF5 е особено популярен в областта на научните изчисления поради способността му за ефективно съхранение и достъп до големи масиви от данни.

	XLSX	CSV	Apache Parquet	HDF5	Pandas DataFrame	
	Storage	Tabular	Tabular	Columnar	Hierarchical	Tabular
	Usage	Office tasks, data presentation	Simple data exchange	Big data, analytics	Scientific data, large volumes	Data analysis, manipulation
	Compression	Built-in	None	High	Built-in	None (in-memory)
	Performance	Low	Medium	High	High	High (memory dependent)
	Complexity	High (formatting, styles)	Low	Medium	Medium	Low
	Data Type Support	Limited	Very limited	Extended	Extended	Extended
	Scalability	Low	Low	High	High	Medium (memory limited)

**Фиг. 8.1-2 Сравнение на форматите за данни, показващо основните разлики в аспектите на съхранение и обработка.**

За да се извърши сравнителен анализ на форматите, използвани на етапа Load на процеса ETL, беше създадена таблица, показваща размерите на файловете и времето за четене (фиг. 8.1-3). В изследването бяха използвани файлове с идентични данни: таблицата съдържаше 10 000 реда и 10 колони, запълнени със случаен стойности.

В проучването са включени следните формати за съхранение: CSV, Parquet, XLSX и HDF5, както и техните компресирани версии в архиви ZIP. Суровите данни са генериирани с помощта на библиотеката NumPy и са представени като структура Pandas DataFrame. Процесът на тестване се състои от следните стъпки:

- Записване на файлове: рамката с данни се записва в четири различни формата: CSV, Parquet, XLSX и HDF5. Всеки формат има уникални характеристики по отношение на начина на съхраняване на данните, което се отразява на размера на файла и скоростта на четене.
- Компресиране на файлове ZIP: за да се анализира ефективността на стандартното компресиране, всеки файл беше допълнително компресиран в архив ZIP.
- Четене на файлове (ETL - Load): времето за четене беше измерено за всеки файл след разопаковането му от ZIP. Това позволява да се оцени скоростта на достъп до данните след изваждането им от архива.

Важно е да се отбележи, че Pandas DataFrame не е използван директно в анализа на размера или

времето за четене, тъй като той не представлява самостоятелен формат за съхранение. Той служи само като междинна структура за генериране и последващо записване на данни в различни формати.



Фиг. 8.1-3 Сравнение на форматите за съхранение по размер и скорост на четене.

Файловете CSV и HDF5 демонстрират (Фигура 8.1-3) висока ефективност на компресиране, като значително намаляват размера си, когато са опаковани в ZIP, което може да бъде особено полезно при сценарии, изискващи оптимизиране на съхранението. Файловете XLSX, от друга страна, практически не подлежат на компресиране и размерът им в ZIP остава сравним с оригинала, което ги прави по-малко благоприятни за използване при големи обеми данни или в среди, където скоростта на достъп до данните е важна. Освен това времето за четене на XLSX е значително по-голямо в сравнение с другите формати, което го прави по-малко предпочитан за бързи операции по четене на данни. Apache Parquet демонстрира висока производителност при аналитични задачи и големи обеми данни благодарение на своята колонарна структура.

## Оптимизиране на съхранението с Apache Parquet

Един от популярните формати за съхраняване и обработка на големи данни е Apache Parquet. Този формат е разработен специално за съхранение на данни в колони (подобно на Pandas), което позволява значително да се намали обемът на паметта и да се увеличи скоростта на аналитичните заявки. За разлика от традиционните формати, като CSV и XLSX, Parquet поддържа собствена компресия и е оптимизиран за системи за големи данни, включително Spark, Hadoop и облачно съхранение.

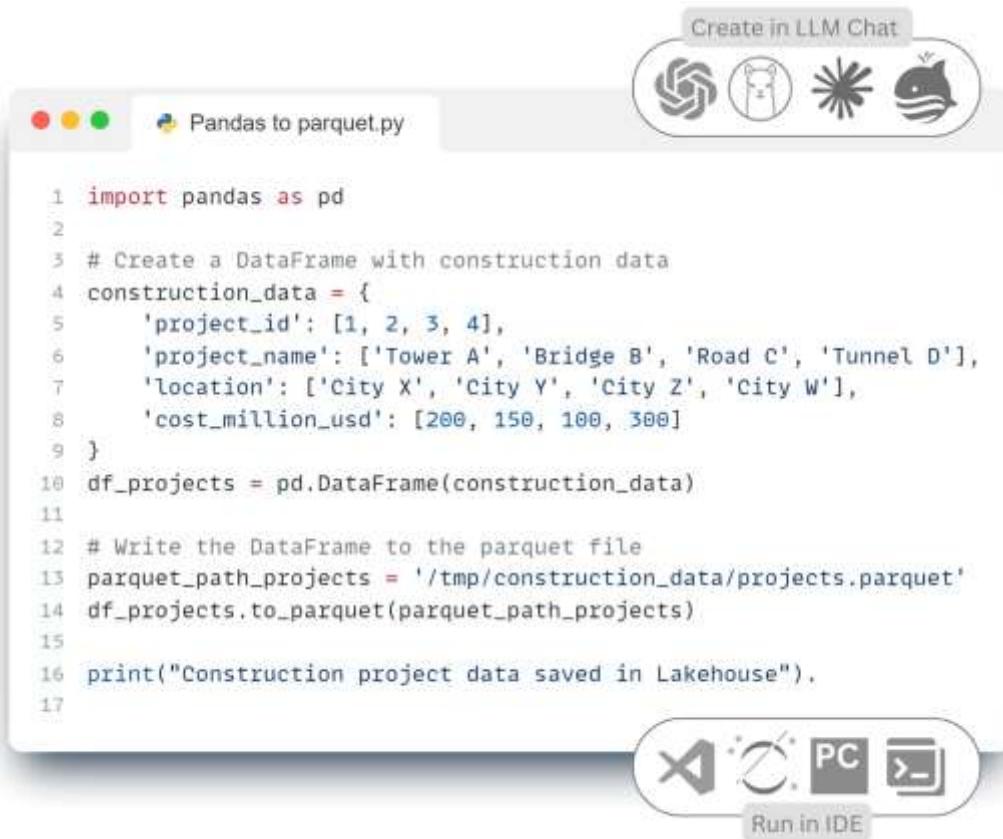
Основните характеристики на Parquet включват поддръжка на компресиране и кодиране на данни, което значително намалява размера на паметта и ускорява операциите по четене на данни, като работи директно с желаните колони, а не с всички редове данни.

За илюстративен пример колко лесно е да се получи необходимият код за преобразуване на данни в Apache Parquet, нека използваме LLM.

- ❷ Из pratете текстова заявка до чата LLM (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN):

Напишете код за записване на данни от Pandas DataFrame в Apache Parquet. ↴

- ❸ Отговор на LLM :



```

1 import pandas as pd
2
3 # Create a DataFrame with construction data
4 construction_data = {
5     'project_id': [1, 2, 3, 4],
6     'project_name': ['Tower A', 'Bridge B', 'Road C', 'Tunnel D'],
7     'location': ['City X', 'City Y', 'City Z', 'City W'],
8     'cost_million_usd': [200, 150, 100, 300]
9 }
10 df_projects = pd.DataFrame(construction_data)
11
12 # Write the DataFrame to the parquet file
13 parquet_path_projects = '/tmp/construction_data/projects.parquet'
14 df_projects.to_parquet(parquet_path_projects)
15
16 print("Construction project data saved in Lakehouse").
17

```

The screenshot shows a Jupyter Notebook cell with the title "Pandas to parquet.py". The cell contains the provided Python code. Above the cell, there is a toolbar with icons for "Create in LLM Chat" and four LLM logos: CHATGP, LlaMa, Mistral DeepSeek, and Grok. Below the cell, there is a "Run in IDE" button.

**Фигура 8.1-4 Прехвърляне на данни за рамки от данни от RAM в ефективен за съхранение формат Apache Parquet с помощта на няколко реда на Python .**

Следващ пример: нека симулираме процеса ETL с данни, съхранявани във формат Parquet, за филтриране на проекти по определена стойност на един от атрибутите "cost\_million\_usd" (фиг. 8.1-4).

- ❷ Докато чатът продължава, изпратете текстова заявка до LLM:

Напишете код, в който искаме да филтрираме данните в таблица и да запазим само тези проекти (редове от таблицата) от данните на Apache Parquet, чиято цена (параметър cost\_million\_usd) надвишава 150 милиона щатски долара. ↴

- ❸ Отговор на LLM :

```

1 # Loading construction project data
2 df_loaded_projects = pd.read_parquet(parquet_path_projects)
3
4 # Filtering out projects with a value greater than $150 million
5 df_filtered_projects = df_loaded_projects[df_loaded_projects['cost_million_usd'] > 150]
6
7 # Saving the filtered data to a new Parquet file
8 filtered_parquet_path_projects = '/tmp/construction_data/filtered_projects.parquet'
9 df_filtered_projects.to_parquet(filtered_parquet_path_projects)
10
11 print("Filtered construction projects data saved to Lakehouse.")

```

Create in LLM Chat

Run in IDE

**Фигура 8.1-5** Процесът ETL при работа с данни във формат Apache Parquet изглежда по същия начин, както при други структурирани формати.

Използването на формата Parquet (в сравнение с XLSX, CSV и т.н.) значително намалява обема на съхраняваната информация и ускорява операциите по търсене. Това го прави отличен както за съхраняване, така и за анализ на данни. Parquet се интегрира с различни системи за обработка, като осигурява ефективен достъп в хибридни архитектури.

Ефективният формат за съхранение обаче е само един от елементите на цялостното преживяване с данните. За да се създаде устойчива и мащабируема среда, е необходима добре проектирана архитектура за управление на данните. Системите от клас DWH (Data Warehouse) изпълняват тази функция. Те осигуряват агрегиране на данни от хетерогенни източници, прозрачност на бизнес процесите и възможност за комплексен анализ с помощта на BI-инструменти и алгоритми за машинно обучение.

## DWH: Data Warehouse складове за данни

Точно както форматът Parquet е оптимизиран за ефективно съхранение на големи обеми информация, така и Data Warehouse е оптимизиран за интегриране и структуриране на данни, за да подпомогне анализите, прогнозирането и вземането на управленски решения.

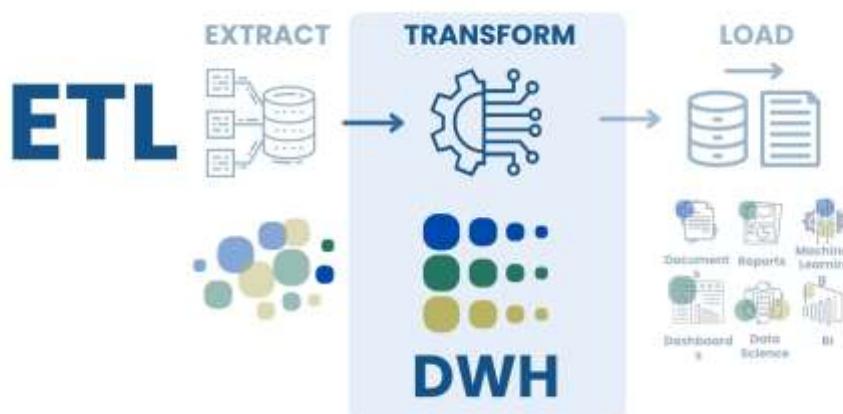
В съвременните компании данните идват от много различни източници: ERP, CAFM, CPM, CRM системи, счетоводство и управление на складове, цифрови CAD модели на сгради, IoT сензори и други решения. За да се получи цялостна картина, не е достатъчно просто да се събират данни - те трябва да бъдат организирани, стандартизираны и централизирани в единно хранилище. Точно това прави DWH - централизирана система за съхранение, която ви позволява да обединявате информация от различни източници, да я структурирате и да я предоставяте за анализ и стратегическо управление.

**DWH (Data Warehouse)** е централизирана система за съхранение на данни, която обединява информация от множество източници, структурира я и я предоставя за анализ и отчитане.

В много компании данните са разпръснати в различни системи, които разглеждахме в първите части на книгата (фигура 1.2-4). DWH интегрира тези източници, като осигурява пълна прозрачност и надеждност на информацията. DWH складът за данни е специализирана база данни (голяма база данни), която събира, обработва и съхранява данни от множество източници. Основните характеристики на DWH са:

- **Използване на ETL -процеси** (Extract, Transform, Load) - извличане на данни от източници, почистването им, трансформирането им, зареждането им в хранилището и автоматизиране на тези процеси, които бяха разгледани в седма част на книгата.
- **Границост на данните** - данните в DWH могат да се съхраняват както в обобщена форма (обобщени отчети), така и в гранулирана форма (необработени данни). От 2024 г. нататък именно доставчиците на CAD- започват да говорят за гранулирани данни [125], което може би показва, че индустрията се подготвя за преход към специализирано съхранение в облак за обработка на данни от цифрови модели на сгради.
- **Поддържането на анализи и прогнозиране** - складовете за данни осигуряват основата за инструментите за BI, анализа на големи данни и машинното обучение.

DWH служи като основа за бизнес разузнаване, което позволява анализ на ключови показатели за ефективност, прогнозиране на продажбите, покупките и разходите, както и автоматизирано отчитане и визуализация на данни (фигура 8.1-6).



**Фиг. 8.1-6** В ETL -процеса DWH може да действа като централно хранилище, където данните, извлечени от различни системи, преминават през етапи на преобразуване и разтоварване.

DWH играе ключова роля в интегрирането, изчистването и структурирането на информацията, осигурявайки солидна основа за бизнес разузнаването и процесите на вземане на решения. Въпреки това в съвременната среда, в която обемите на данните нарастват бързо, а източниците на данни стават все по-разнообразни, традиционният DWH подход за съхранение на информация често изиска разширяване под формата на ELT и Data Lake.

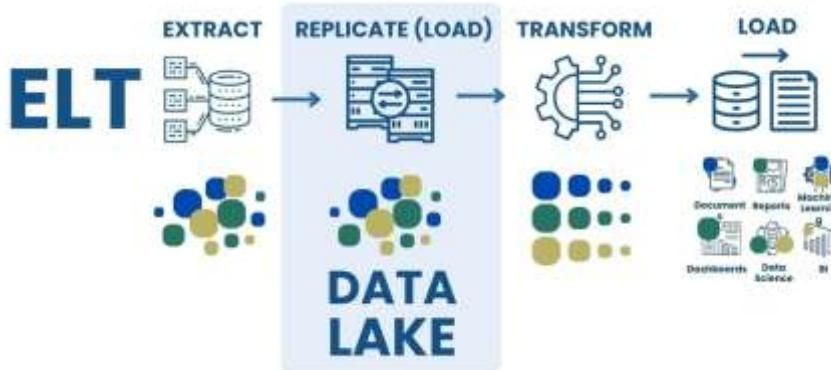
### **Езеро от данни - еволюция на ETL към ELT: от традиционно почистване към гъвкава обработка**

Класически DWH - складове за данни, предназначени за съхраняване на структурирани данни във формат, оптимизиран за аналитични заявки, са се сблъскали с ограничения при обработката на неструктурирани данни и мащабируемостта. В отговор на тези предизвикателства се появила езерата от данни), които предлагат гъвкаво съхранение на големи количества хетерогенни данни.

**Data Lake** предлага алтернативен DWH -подход, който ви позволява да работите с неструктурни, полуструктурни и необработени данни без предварителна твърда схема. Този метод на съхранение често е от значение за обработката на данни в реално време, машинното обучение и разширени анализи. За разлика от DWH, който структурира и агрегира данните преди зареждане, Data Lake позволява информацията да се съхранява в сыръв вид, като по този начин осигурява гъвкавост и мащабируемост.

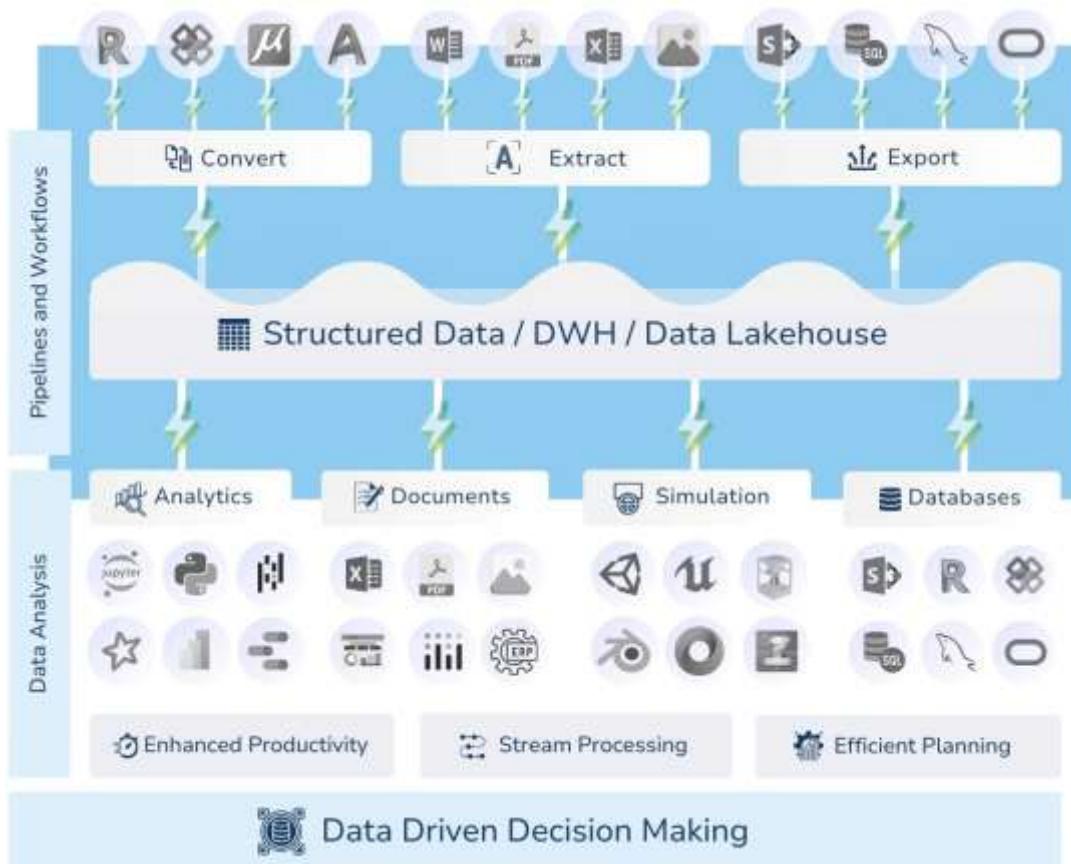
Именно разочароването от традиционните складове за данни (СУБД, DWH) и интересът към "големите данни" доведоха до появата на езерата за данни, при които вместо сложния ETL данните вече просто се качват в слабо структурирано хранилище, а обработката се извършва на етапа на анализ:

- В традиционните складове за данни данните обикновено се подлагат на предварителна обработка, преобразуване и почистване (ETL - Extract, Transform, Load), преди да бъдат заредени в склада (фигура 8.1-6). Това означава, че данните са структурирани и оптимизирани за конкретни бъдещи задачи за анализ и отчитане. Акцентът е върху поддържането на висока производителност на заявките и интегритет на данните. Този подход обаче може да бъде скъп и по-малко гъвкав по отношение на интегрирането на нови типове данни и бързо променящи се схеми на данните.
- От друга страна, езерата от данни са предназначени за съхраняване на големи количества необработени данни в оригиналния им формат (Фигура 8.1-7). Процесът ETL (Извличане, Трансформиране, Зареждане), при който данните първо се зареждат в хранилището "във вида, в който са", и едва след това могат да бъдат трансформирани и анализирани според нуждите. Това осигурява по-голяма гъвкавост и възможност за съхраняване на хетерогенни данни, включително неструктурирани данни като текст, изображения и логове.



Фиг. 8.1-7 За разлика от ETL, езерото от данни използва ELT, при което информацията първо се качва в "сиров" вид и преобразуването се извършва на етапа на качване.

Традиционните складове за данни се фокусират върху предварителната обработка на данните, за да осигурят висока производителност на заявките, докато езерата за данни дават приоритет на гъвкавостта: те съхраняват необработени данни и ги трансформират според нуждите (фигура 8.1-8).



**Фиг. 8.1-8 Съвременните концепции за съхранение имат за цел да съхраняват и обработват всички видове данни за целите на вземането на решения.**

Въпреки всички предимства обаче езерата от данни не са лишени от недостатъци. Липсата на строга структура и сложността на управлението на информацията могат да доведат до хаос, при който данните се дублират, противоречат си или стават нерелевантни. Освен това търсенето и анализирането на данни в такова хранилище изисква значителни усилия, особено когато става въпрос за разнородна информация. За да се преодолеят тези ограничения и да се съчетаят най-добрите характеристики на традиционните складове за данни и езерата от данни, е разработена архитектурата Data Lakehouse.

### **Архитектура на езерото за данни: синергия на складове и езера за данни**

За да се съчетаят най-добрите характеристики на DWH (структурен, управляем, високопроизводителен анализ) и Data Lake (машабируемост, работа с хетерогенни данни), беше разработен подходът Data Lakehouse. Тази архитектура съчетава гъвкавостта на езерата от данни с мощните инструменти за обработка и управление, характерни за традиционните складове, като постига баланс между съхранението, анализа и машинното обучение. Data Lakehouse е синтез на езера от данни и складове за данни, като съчетава гъвкавостта и машабируемостта на първите с управляемостта и оптимизацията на заявките на вторите.

**Data Lakehouse** е архитектурен подход, който има за цел да съчетае гъвкавостта и мащабируемостта на езерата от данни с управляемостта и производителността на заявките на складовете за данни (фигура 8.1-9).

Основните характеристики на Data Lakehouse включват:

- **Отворен формат за съхранение на данни:** използването на отворени формати за съхранение на данни, като например Apache Parquet, осигурява ефективност и оптимизирани заявки.
- **Схема само за четене:** за разлика от традиционния подход на схема само за запис в DWH, Lakehouse поддържа схема само за четене, която позволява по-голяма гъвкавост при управлението на структурата на данните.
- **Гъвкавост и мащабируемост:** поддържа съхранение и анализ на структурирани и неструктурни данни, като осигурява висока производителност на заявките чрез оптимизация на ниво съхранение.

Data Lakehouse предлага компромисно решение, което съчетава предимствата и на двета подхода, което го прави идеално за съвременни аналитични натоварвания, изискващи гъвкавост при обработката на данни.



Фиг. 8.1-9 Data Lakehouse е следващото поколение системи за съхранение на данни, предназначени да отговарят на сложни и постоянно променящи се изисквания.

Идеята на съвременните складове за данни изглежда прости: ако всички данни са на едно място, е по-лесно да се анализират. На практика обаче всичко не е толкова гладко. Представете си, че дадено дружество реши напълно да изостави обичайните системи за счетоводство и управление (ERP, PMIS, CAFM или други), като ги замени с едно огромно езеро от данни, до което всички имат достъп. Какво ще се случи? Най-вероятно ще настъпи хаос: данните ще се дублират, ще са противоречиви, а критичната информация ще бъде загубена или повредена. Дори ако езерото от данни се използва само за анализи, без правилно управление то ще бъде сериозно компрометирано:

- Данните са трудни за разбиране: в конвенционалните системи данните имат ясна структура, но в езерото те са просто огромно струпване на файлове и таблици. За да намери нещо, трябва да разбере за какво отговаря всеки ред и колона.
- Данните могат да бъдат неточни: ако на едно място се съхраняват много версии на една и съща информация, е трудно да се разбере коя версия е актуална. В резултат на това решенията се вземат въз основа на неактуални или погрешни данни.
- Трудно е да се подгответят данни за работа: данните трябва не само да се съхраняват, но и да се представят в удобна форма - под формата на отчети, графики, таблици. В традиционните системи това се прави автоматично, но в езерата от данни е необходима допълнителна обработка.

В резултат на това всяка концепция за съхранение на данни има свои собствени характеристики, подходи за обработка и бизнес приложения. Традиционните бази данни се фокусират върху транзакционни операции, складовете за данни (DWH) осигуряват структура за анализи, езерата от данни (Data Lake) съхраняват информация в необработен вид, а хиbridните складове (Data Lakehouse) съчетават предимствата на DWH и Data Lake (фигура 8.1-10).

	Traditional Approach	Data Warehouse	Data Lake	Data Lakehouse
Data Types	Relational Databases	Structured, ready for analytics	Raw, semi-structured, or unstructured	Mix of structured and unstructured
Use Cases	Transactional Systems	Reporting, dashboards, BI	Big data storage, AI, advanced analytics	Hybrid analytics, AI, real-time data
Processing	OLTP – real-time transactions	ETL – clean and structure before analysis	ELT – store raw data, transform later	ELT with optimized storage and real-time processing
Storage	On-premise servers	Centralized, SQL-based	Decentralized, flexible formats	Combines advantages of DWH and DL
Common Tools	MySQL, PostgreSQL	Snowflake, Redshift, BigQuery	Hadoop, AWS S3, Azure Data Lake	Databricks, Snowflake, Google BigLake

**Фиг. 8.1-10 DWH, Data Lake и Data Lakehouse: основни разлики във видовете данни, сценарийите на използване, методите за обработка и подходите за съхранение.**

Изборът на архитектура за съхранение е сложен процес, който зависи от бизнес нуждите, обема на информацията и изискванията за анализ. Всяко решение има своите плусове и минуси: DWH осигурява структура, Data Lake осигурява гъвкавост, а Lakehouse осигурява баланс между двете. Организациите рядко са ограничени до една-единствена архитектура на данните.

Независимо от избраната архитектура, автоматизираните системи за управление на данни значително превъзхождат ръчните методи. Те свеждат до минимум човешките грешки, ускоряват обработката на информацията и осигуряват прозрачност и проследимост на данните на всички етапи на бизнес процесите.

И докато централизираните складове за данни вече са превърнали в индустриски стандарт в много области на икономиката, ситуацията в строителството остава фрагментирана. Данните тук

са разпределени между различни платформи (CDE, PMIS, ERP и др.), което затруднява създаването на единна картина на случващото се и изисква архитектури, способни да комбинират тези източници в съгласувана, използваема за анализи цифрова среда.

## CDE, PMIS, ERP или DWH и езеро от данни

Някои строителни и инженерингови компании вече използват концепцията за обща среда за данни (CDE) в съответствие с ISO 19650. По същество CDE изпълнява същите функции като склада за данни (DWH) в други отрасли: централизиране на информацията, осигуряване на контрол на версии, осигуряване на достъп до потвърдена информация.

**Общата среда за данни (CDE)** е централизирано цифрово пространство, което се използва за управление, съхранение, споделяне и сътрудничество по отношение на информацията за проекта през всички фази от жизнения цикъл на обекта. CDE често се внедрява с помощта на облачни технологии и се интегрира с CAD (BIM) системи.

Финансовият, търговският, логистичният и промишленият сектор от десетилетия използват централизирани системи за управление на данни, които комбинират информация от различни източници, контролират нейната значимост и предоставят анализи. CDE доразвива тези принципи, като ги адаптира към предизвикателствата на проектирането на сгради и управлението на жизнения им цикъл.

Подобно на DWH, CDE структурира данните, улавя промените и осигурява единна точка за достъп до проверената информация. С преминаването към облака и интеграцията с аналитични инструменти разликите между тях стават все по-малко очевидни. Като се добавят към CDE гранулираните данни, чиято концепция се обсъжда от CAD -доставчиците от 2023 г.[93, 125], могат да се видят още повече паралели с класическия DWH.

По-рано в главата "Строителни ERP и PMIS системи" вече разглеждахме PMIS (Информационна система за управление на проекти) и ERP (Планиране на ресурсите на предприятието). В строителните проекти CDE и PMIS работят заедно: CDE служи като ханилище за данни, включително чертежи, модели и проектна документация, докато PMIS управлява процеси като контрол на срокове, задачи, ресурси и бюджети.

ERP, отговаряща за управлението на бизнеса като цяло (финанси, снабдяване, персонал, производство), може да се интегрира с PMIS, осигуряваща контрол на разходите и бюджета на ниво компания. За анализ и отчитане DWH може да се използва за събиране, структуриране и обобщаване на данни от CDE, PMIS и ERP, за да се оценят финансовите ключови показатели за ефективност (ROI) и да се идентифицират модели. От своя страна, езерото от данни (DL) може да допълни DWH, като съхранява необработени и неструктурирани данни (напр. логове, данни от сензори, изображения). Тези данни могат да бъдат обработени и заредени в DWH за по-нататъшен анализ.

Така CDE и PMIS се фокусират върху управлението на проекти, ERP - върху бизнес процесите, а DWH и Data Lake - върху анализите и операциите с данни.

При сравняване на системите CDE, PMIS и ERP с DWH и Data Lake се наблюдават значителни разлики по отношение на независимостта от доставчика, разходите, гъвкавостта на интеграцията, независимостта на данните, скоростта на адаптиране към промените и аналитичните възможности (фигура 8.1-11). Традиционните системи, като CDE, PMIS и ERP, често са обвързани с конкретни решения и стандарти на доставчиците, което ги прави по-малко гъвкави и увеличава разходите им поради лицензи и поддръжка. Освен това данните в такива системи често са капсулирани в патентовани, затворени формати, което ограничава тяхното използване и анализ.

		CDE, PMIS, ERP	DWH, Data Lake
	Vendor Dependency	High (tied to specific solutions and standards of vendors)	Low (flexibility in tool and platform choice)
	Integration Flexibility	Limited (integration depends on vendor solutions)	High (easily integrates with various data sources)
	Cost	High (licensing and support costs)	Relatively lower (use of open technologies and platforms)
	Data Independence	Low (data often locked in proprietary formats)	High (data stored in open and accessible formats)
	Adaptability to Changes	Slow (changes require vendor approval and integration)	Fast (adaptation and data structure modification without intermediaries)
	Analytical Capabilities	Limited (dependent on vendor-provided solutions)	Extensive (support for a wide range of analytical tools)

Фиг. 8.1-11 DWH и езерото от данни предлагат по-голяма гъвкавост и независимост на данните, отколкото системи като CDE, PMIS и ERP.

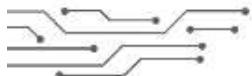
За разлика от тях DWH и Data Lake осигуряват по-голяма гъвкавост при интегрирането с различни източници на данни, а използването на отворени технологии и платформи помага за намаляване на общите разходи за притежание. Освен това DWH и Data Lake поддържат широк набор от аналитични инструменти, което подобрява възможностите за анализ и управление.

С развитието на инструментите за обратен инженеринг на CAD форматите и достъпа до базите данни на CAD приложенията въпросът става все по-актуален: доколко е оправдано да продължаваме да използваме затворени, изолирани платформи, ако проектните данни трябва да бъдат достъпни за широк кръг специалисти, работещи в десетки изпълнители и проектантски организации?

Тази технологична зависимост от конкретен доставчик може значително да ограничи гъвкавостта на управлението на данните, да забави реакциите при промени в проекта и да възпрепятства ефективното сътрудничество между участниците.

Традиционните подходи към управлението на данни - включително DWH, Data Lake, CDE и PMIS - са насочени предимно към съхраняване, структуриране и обработка на информация. Въпреки това, с развитието на изкуствения интелект и машинното обучение, нараства необходимостта от нови начини за организиране на данните, които не само обобщават, но и разкриват сложни връзки, намират скрити модели и осигуряват незабавен достъп до най-подходящата информация.

Специална роля в тази насока започват да играят векторните бази данни - нов тип хранилище, оптимизирано за високоразмерни вграждания.



## ГЛАВА 8.2.

# УПРАВЛЕНИЕ НА ХРАНИЛИЩА ЗА ДАННИ И ПРЕДОТВРАТИВАНЕ НА ХАОС

### Векторни бази данни и гранична кутия

Векторните бази данни са нов клас хранилища, които не просто съхраняват данни, а позволяват търсене по смисъл, сравняване на обекти по семантична близост и създаване на интелигентни системи: от препоръки до автоматичен анализ и генериране на контекст. За разлика от традиционните бази данни, които се фокусират върху точните съвпадения, векторните бази данни намират сходни обекти въз основа на атрибути - дори ако няма точно съвпадение

**Векторна база данни** е специализиран тип база данни, която съхранява данни като многомерни вектори, всеки от които представя определени характеристики или качества. Тези вектори могат да имат различен брой измерения в зависимост от сложността на данните (в един случай това могат да бъдат няколко измерения, а в друг — хиляди).

Основното предимство на векторните бази данни е търсенето по семантична значимост, а не по точно съвпадение на стойности. Вместо SQL- и Pandas-запитвания с филтри "equals" или "contains" се използва търсенето на най-близките съседи (k-NN) (ще говорим повече за k-NN в следващата част на книгата) в пространството на признаките.

С разработването на LLM (Large Language Models) и генеративни модели взаимодействието с базите данни започва да се променя. Вече е възможно да се правят запитвания за данни на естествен език, да се получават семантични търсения по документи, автоматично да се извлечат ключови термини и да се изграждат контекстуални връзки между обектите - и всичко това без да е необходимо да се владее SQL или да се познава структурата на таблиците. Това беше разгледано по-подробно в раздела "МЛП и тяхната роля в обработката на данни и бизнес процесите".

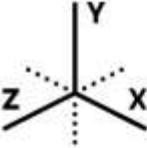
Важно е обаче да се разбере, че LLM не структурират и не подреждат информацията автоматично. Моделът просто преминава през данните и намира най-подходящата част от тях въз основа на контекста на заявката. Ако данните не са били предварително почистени или трансформирани, дълбокото търсене ще бъде като опит за намиране на отговор в цифров "боклук" - може и да работи, но качеството на резултатите ще бъде по-ниско. В идеалния случай, ако данните могат да бъдат структурирани (напр. да се преведат документите в Markdown) и заредени във векторна база данни. Това значително повишава точността и релевантността на резултата.

Първоначално векторните бази данни се използваха в машинното обучение, но днес намират все повече приложения извън него - в търсачките, персонализирането на съдържанието и интелигентния анализ.

Един от най-очевидните примери за векторния подход в конструирането е Bounding Box (гранич

паралелепипед). Това е геометрична конструкция, описваща границите на даден обект в триизмерното пространство. Bounding Box се определя от минималните и максималните координати X, Y и Z, като образува "кутия" около обекта. Този метод ви позволява да оцените размера и разположението на даден елемент, без да се налага да анализирате цялата геометрия.

Всеки Bounding Box може да бъде представен като вектор в многомерно пространство: например [x, y, z, width, height, depth] - вече 6 измерения (фиг. 8.2-1)



### Bounding Box

	minX	maxX	minY	maxY	minZ	maxZ	Width	Height	Depth
Column	-15	-5	-25	-15	0	10	10	10	20
Stairs	-5	5	-15	-5	0	10	10	10	10
Door	5	15	5	15	0	10	10	10	10
Window	25	35	-35	-25	10	30	10	20	20
Balcony	15	25	-5	5	20	40	10	20	20

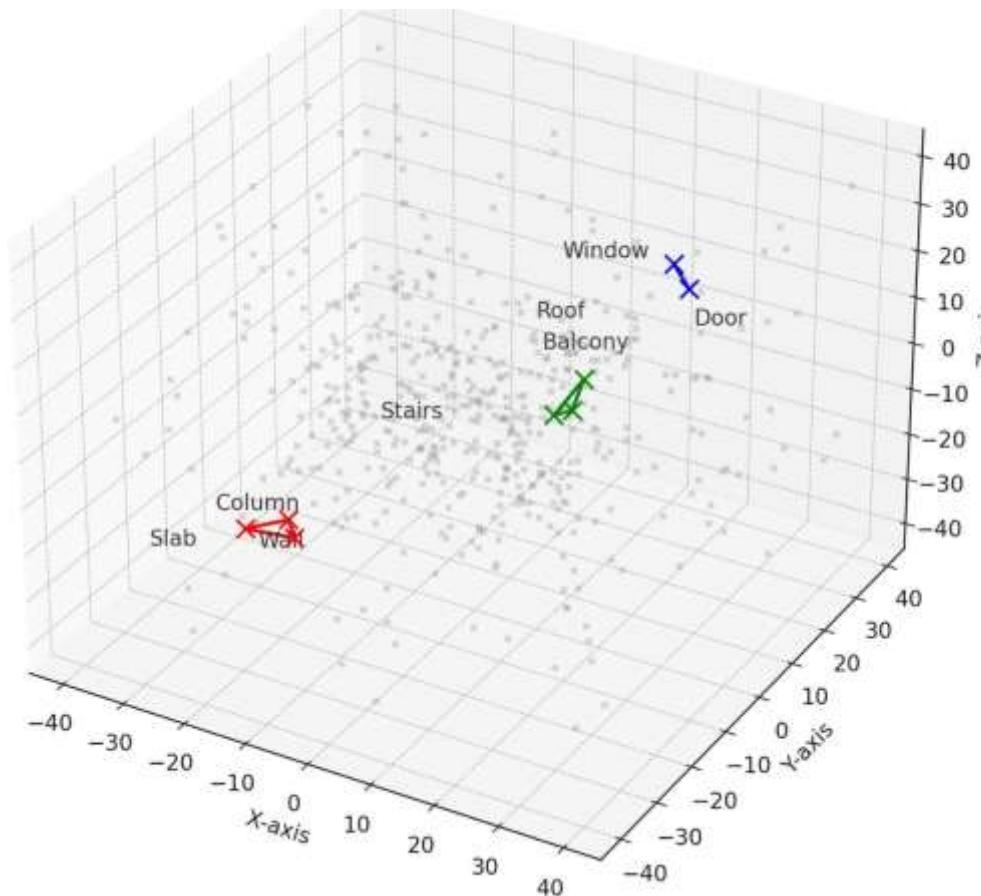
Фиг. 8.2-1 Bounding Box - информацията за координатите на елементите и тяхното местоположение в модела на проекта е аналогична на векторна база данни.

Това представяне на данни улеснява много задачи, включително проверка за пресичания между обекти, планиране на пространственото разпределение на елементите на сградата и извършване на автоматизирани изчисления. Bounding Box може да служи като мост между сложните 3D модели и традиционните векторни бази данни, като ви позволява ефективно да използвате предимствата и на двата подхода в архитектурното и инженерното моделиране

Bounding Box е "векторизиране на геометрията", а embedding (начин за трансформиране на нещо абстрактно) е "векторизиране на смисъла". И двата подхода ви позволяват да преминете от ръчно търсене към интелигентно търсене, независимо дали става въпрос за 3D-обекти в модел на проект или понятия в текст.

Търсенето на обекти в проекта (например "намери всички прозорци с ширина > 1,5 м") е подобно на търсенето на най-близки съседи (k-NN) във векторна база данни, където критериите определят "зона" в пространството на признаките. (ще говорим повече за търсенето на най-близки съседи k-NN в следващата част за машинното обучение) (фиг. 8.2-2). Ако към атриутите на ограничителната зона добавим допълнителни параметри (материал, тегло, време за производство), таблицата се превръща във високоразмерен вектор, където всеки атрибут е ново измерение. Това е по-близо до

съвременните векторни бази, където измеренията се броят в стотици или хиляди (например вграждане от невронни мрежи).



Фиг. 8.2-2 Търсене на обекти в проекта с помощта на векторни бази данни.

Подходът, използван в Bounding Box, е приложим не само за геометрични обекти, но и за анализ на текст и език. Векторните представяния на данни вече се използват активно в обработката на естествен език (NLP). Точно както обектите в един строителен проект могат да бъдат групирани според тяхната пространствена близост (фигура 8.2-2), така и думите в един текст могат да бъдат анализирани според тяхната семантична и контекстуална близост.

Например думите "архитект", "строителство", "дизайн" ще бъдат една до друга във векторното пространство, защото имат сходно значение. В LLM този механизъм позволява автоматична, без да е необходима ръчна категоризация:

- Определяне на темата на даден текст
- Извършване на семантични търсения в съдържанието на документи
- Генериране на автоматични анотации и резюмета на текст
- Намиране на синоними и свързани термини

Векторните бази данни ви позволяват да анализирате текст и да намирате свързани термини в него по същия начин, по който Bounding Box ви помага да анализирате пространствени обекти в 3D -модели. Примерът с Bounding Box на елементи от проекта помага да се разбере, че векторното представяне не е чисто "изкуствена" концепция от ML, а естествен начин за структуриране на данни за решаване на приложни задачи, независимо дали става въпрос за търсене на колони в CAD проект или за семантично близки изображения в база данни.

Специалистите, работещи с бази данни, трябва да обърнат внимание на векторните хранилища. Тяхното разпространение показва нов етап в развитието на базите данни, при който класическите релационни системи и технологиите, ориентирани към изкуствения интелект, започват да се преплитат, формирајки хибридни решения на бъдещето.

Потребителите, които разработват сложни и широкомащабни приложения с изкуствен интелект, ще използват специализирани бази данни за векторно търсене. В същото време тези, които се нуждаят само от отделни AI-функции за интегриране в съществуващи приложения, по-скоро ще изберат вградени възможности за векторно търсене в базите данни, които вече използват (PostgreSQL, Redis).

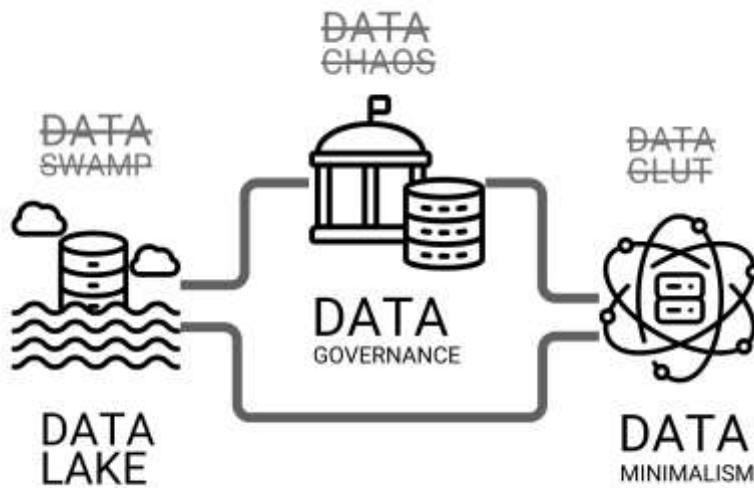
Въпреки че системи като DWH, Data Lake, CDE, PMIS, векторни бази данни и други предлагат различни подходи за съхранение и управление на данни, тяхната ефективност се определя не само от архитектурата им, но и от това колко добре са организирани и управлявани самите данни. Дори когато се използват съвременни решения - било то векторни бази данни, класически релационни СУБД или складове от типа Data Lake - липсата на ясни правила за управление, структуриране и актуализиране на данните може да доведе до същите трудности, с които се сблъскват потребителите, работещи с разнородни файлове и данни в различни формати.

Без управление на данните) дори и най-мощните решения могат да станат хаотични и неструктурирани, превръщайки езерата от данни в блата от данни). За да избегнат това, компаниите трябва не само да изберат правилната архитектура за съхранение, но и да приложат стратегии за минимализиране на данните), управление на достъпа и контрол на качеството, за да превърнат данните в ефективен инструмент за вземане на решения.

## **Управление на данните), минимализъм на данните) и блато на данните)**

Разбирането и прилагането на концепциите за управление на данните), минимализъм на данните) и предотвратяване на блатото от данни) са ключови за успешното управление на хранилищата на данни и осигуряването на бизнес стойност (фигура 8.2-3).

Според проучване на Gartner (2017 г.) 85% от проектите за големи данни се провалят, а една от основните причини е недостатъчното качество на данните и управлението на данните [144].



Фиг. 8.2-3 Някои от ключовите аспекти на управлението на данните са управлението на данните и минимализъмът на данните.

Управлението на данните (Data Governance) е основен компонент на управлението на данните, който гарантира, че данните се използват правилно и ефективно във всички бизнес процеси. Става дума не само за установяване на правила и процедури, но и за осигуряване на наличност, надеждност и сигурност на данните:

- Дефиниране и класифициране на данни: ясното дефиниране и класифициране на единиците позволява на организацията да разберат какви единици са необходими в компанията и да определят как да бъдат използвани.
- Права за достъп и управление: разработването на политики и процедури за достъп и управление на данните гарантира, че само упълномощени потребители имат достъп до определени данни.
- Защита на данните от външни заплахи: Защитата на данните от външни заплахи е ключов аспект на управлението на данните. Това включва не само технически мерки, но и обучение на служителите в основите на информационната сигурност.

Минимализъмът на данните (Data Minimalism) е подход за редуциране на данните до най-ценните и значими атрибути и единици във формацията (фигура 8.2-4), като по този начин се намаляват разходите и се подобрява използването на данните:

- Опростяване на процеса на вземане на решения: намаляването на броя на обектите и техните атрибути до най-важните от тях опростява процеса на вземане на решения, като намалява времето и ресурсите, необходими за анализ и обработка на данните.
- Фокусиране върху важните неща: Избирането на най-подходящите обекти и атрибути ви позволява да се съсредоточите върху информацията, която наистина е важна за бизнеса, като елиминирате шума и ненужните данни.
- Ефективно разпределение на ресурсите: минимизирането на данните позволява по-ефективно разпределение на ресурсите, намаляване на разходите за съхранение и обработка на данни, подобряване на качеството и сигурността на данните.

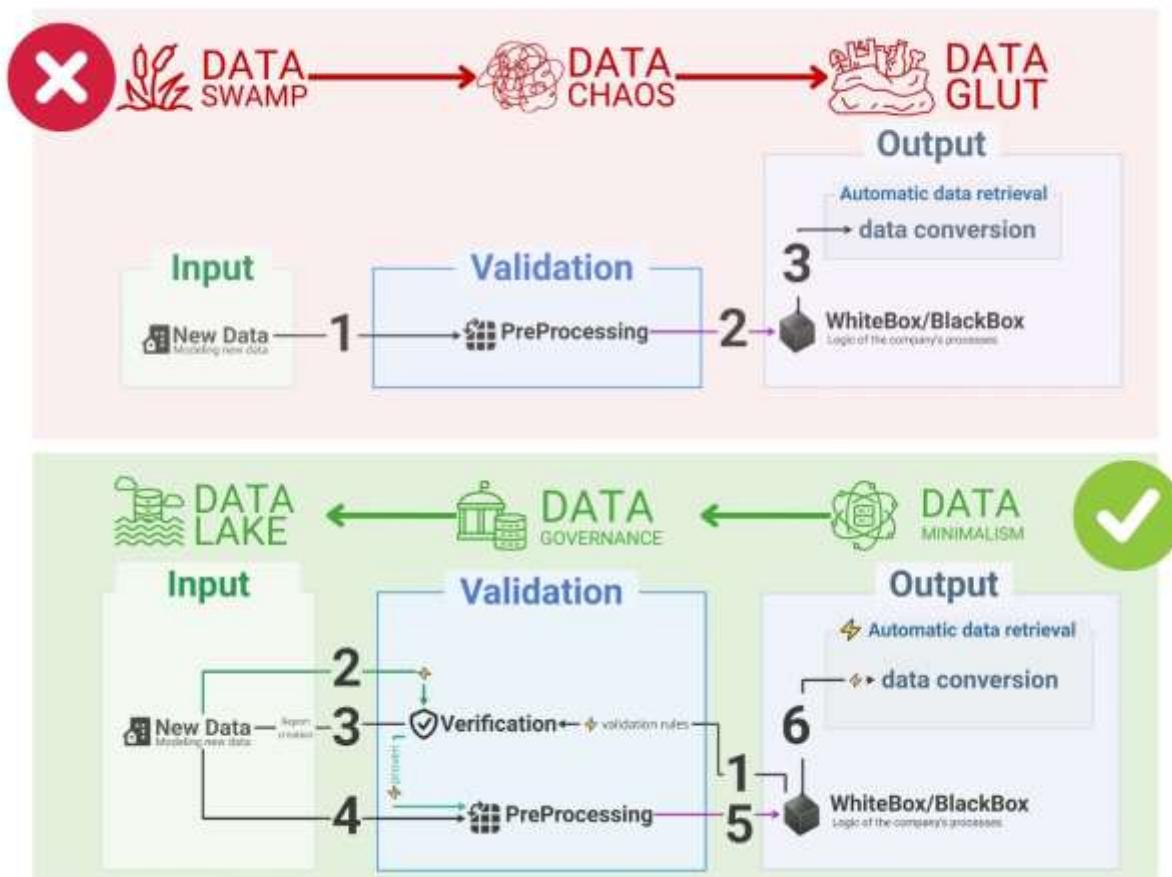
Логиката на работа с данни не трябва да започва със самото им създаване (фиг. 8.2-4), а с разбирането на бъдещите сценарии за използване на тези данни още преди да започне процесът на генериране. Този подход позволява предварително да се определят минимално необходимите изисквания към атрибутите, техните типове и гранични стойности. Тези изисквания формират основата за създаване на коректни и стабилни същности в информационния модел. Предварителното разбиране на целите и употребите на данните допринася за формирането на структура, подходяща за анализ. Повече подробности за подходите към моделирането на данни на концептуално, логическо и физическо ниво бяха разгледани в главата "Моделиране на данни: концептуален, логически и физически модел".

В традиционните бизнес процеси на строителните компании обработката на данни по-често прилича на изхвърляне на данни в блато, където данните първо се създават, а след това специалистите се опитват да ги интегрират в други системи и инструменти.

Блатото от данни (Data Swamp) е резултат от неконтролирано събиране и съхраняване на данни без подходяща организация, структуриране и управление, в резултат на което данните са неструктурirани, трудни за използване и с ниска стойност.

Как да предотвратим превръщането на информационния поток в тресавище:

- **Управление на структурата на данните:** осигуряването на структуриране и категоризиране на данните на помага да се предотврати натрупването на данни, като ги прави подредени и лесно достъпни.
- **Разбиране и тълкуване на данните:** ясното описание на произхода, модификациите и значението на данните гарантира, че те се разбират и тълкуват правилно.
- **Поддържане на качеството на данните:** Редовната поддръжка и почистване на данните помага за поддържане на качеството на данните, тяхната релевантност и стойност за анализите и бизнес процесите.



Фигура 8.2-4 За да избегнете безпорядъка в хранилището на данни, трябва да започнете процеса на създаване на данни, като съберете изисквания за атрибути.

Чрез интегриране на принципите на управление на данните и минимализъм на данните в процесите на управление на данните и активно предотвратяване на превръщането на складовете за данни в блато от данни, организациите могат да увеличат максимално потенциала на своите данни.

Следващият етап в еволюцията на работата с данни, след решаването на проблемите с управлението и минимализма, е стандартизирането на автоматичната обработка, осигуряването на качеството и прилагането на методи, които правят данните използваеми за анализ, преобразуване и вземане на решения. Именно това правят методологията DataOps и VectorOps, които се превръщат във важни инструменти за компаниите, работещи с големи обеми данни и машинно обучение.

## DataOps и VectorOps: нови стандарти за данни

Докато Data Governance отговаря за контрола и организирането на данните, DataOps помага да се гарантира тяхната точност, последователност и гладко протичане в рамките на компанията. Това е особено важно за редица бизнес случаи в строителството, където данните се генерират непрекъснато и изискват навременна обработка. Например в ситуации, в които информационните

модели на сградата, изискванията на проекта и аналитичните отчети трябва да бъдат синхронизирани между различни системи в рамките на един работен ден, ролята на DataOps може да бъде ключова. Тя позволява да се изградят стабилни и повтарящи се процеси за обработка на данни, като се намалява рисъкът от забавяне и загуба на релевантност на информацията.

Управлението на данните само по себе си не е достатъчно - от съществено значение е данните не само да се съхраняват, но и да се използват активно в ежедневните операции. Именно тук се появява DataOps - методология, фокусирана върху автоматизацията, интеграцията и непрекъснатия поток от данни.

**DataOps** се фокусира върху подобряване на сътрудничеството, интеграцията и автоматизацията на потоците от данни в организацията. Възприемането на практиките на DataOps настъпчава точността, последователността и наличността на данните, което е от решаващо значение за приложениета, ориентирани към данните.

Ключови инструменти в екосистемата DataOps са Apache Airflow (Фигура 7.4-4) за оркестрация на работни потоци и Apache NiFi (Фигура 7.4-5) за маршрутизиране и трансформиране на потоци от данни. Заедно тези технологии дават възможност за гъвкави, надеждни и мащабиращи тръбопроводи за данни за автоматична обработка, контрол и интегриране на информация между системите (повече подробности в главата "Автоматичен ETL-конвейер"). При прилагането на подхода DataOps в строителните процеси е важно да се вземат предвид четири основни аспекта:

- Хората и инструментите са по-важни от данните:** изолирани хранилища на данни може да се разглежда като основен проблем, но реалността е по-сложна. В допълнение към фрагментацията на данните, изолацията на екипите и различните инструменти, които те използват, играят важна роля. В строителството с данните работят специалисти от различни дисциплини: инженери и анализатори на данни, екипи по BI и визуализация, както и експерти по управление на проекти и качество. Всеки от тях има различни начини на работа, така че става важно да се създаде екосистема, в която данните да текат свободно между участниците, осигурявайки единна, последователна версия на информацията.
- Автоматизиране на тестването и откриването на грешки:** Строителните данни винаги съдържат грешки, независимо дали става въпрос за неточности в моделите, грешки в изчисленията или остарели спецификации. Редовното тестване на данните и отстраняването на повтарящите се грешки може значително да подобри качеството на данните. Като част от DataOps трябва да въведете автоматизирани механизми за контрол и валидиране, които да следят за коректността на данните, да анализират грешките и да идентифицират моделите, както и да улавят и отстраняват системните грешки във всеки работен процес. Колкото по-висока е степента на автоматизирано валидиране, толкова по-високо е цялостното качество на данните и толкова по-малка е вероятността от грешки в крайните етапи.
- Данните трябва да бъдат тествани по същия начин, както програмния код:** повечето приложения в сградите се основават на обработката на данни, но контролът върху тях често се оставя на второстепенни роли. Ако моделите за машинно обучение са обучени върху неточни данни, това води до неправилни прогнози и финансови загуби. В рамките на DataOps данните трябва да се подлагат на същия строг контрол като програмния код: логически проверки, стрес тестове и оценка на поведението на моделите при промяна на входните стойности. Само валидирани и надеждни данни могат да се използват като основа за управлениски решения.

**4. Наблюдаемост на данните без компромис с производителността:** наблюдението на данните не е просто набор от показатели, а стратегически инструмент за управление на качеството. За да може DataOps да работи ефективно, наблюдаемостта трябва да бъде вградена на всички етапи на работа с данни - от проектирането до експлоатацията. В същото време е важно наблюдението да не забавя системата. В строителните проекти е изключително важно не само да се събират данни, но и да се прави това по такъв начин, че работата на специалистите (напр. проектантите), които създават данните, да не се нарушава по никакъв начин. Този баланс ви позволява да контролирате качеството на данните, без да правите компромис с производителността.

DataOps не е допълнителна тежест за учените, занимаващи се с данни, а гръбнакът на тяхната работа. Чрез внедряването на DataOps строителните компании могат да преминат от хаотично управление на данни към ефективна екосистема, в която данните работят за бизнеса.

На свой ред VectorOps представлява следващ етап от развитието на DataOps, насочен към обработката, съхранението и анализа на многомерни векторни данни (които бяха разгледани в предишната глава). Това е от особено значение в области като цифровите близнази, моделите на невронните мрежи и семантичното търсене, които започват да навлизат в строителната индустрия. VectorOps разчита на векторни бази данни за ефективно съхранение, индексиране и търсене на многомерни представления на обекти.

**VectorOps** е следващата стъпка след DataOps, насочена към обработката, анализа и използването на векторни данни в строителството. За разлика от DataOps, който се фокусира върху потока, последователността и качеството на данните, VectorOps се фокусира върху управлението на многомерните представления на обекти, необходими за машинното обучение.

За разлика от традиционните подходи VectorOps ви позволява да постигнете по-точни описания на обектите, което е от решаващо значение за цифровите близнази, системите за генеративен дизайн и автоматичното откриване на грешки в CAD данните, преобразувани във векторен формат. Комбинираното прилагане на DataOps и VectorOps формира солидна основа за мащабируема, автоматизирана работа с големи обеми информация - от класически таблици до семантично богати пространствени модели.

## Следващи стъпки: от хаотично съхранение към структурирано съхранение

Традиционните подходи за изграждане на хранилища за данни често водят до създаването на различни "информационни силози", в които важните прозрения са недостъпни за анализ и вземане на решения. Съвременните концепции за съхранение на данни, като например Data Warehouse, Data Lake и техните хибриди, дават възможност за обединяване на разнородна информация и централизираното ѝ предоставяне за стрийминг на данни и бизнес разузнаване. Важно е не само да се избере правилната архитектура за съхранение, но и да се приложат управлението на данните (минимализъм на данните), за да се предотврати превръщането на съоръженията за съхранение в неконтролируеми блата от данни).

За да обобщим тази част, си струва да подчертаем основните практически стъпки, които ще ви

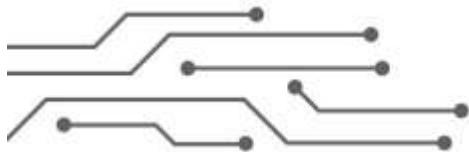
помогнат да приложите разгледаните концепции към ежедневните си задачи:

- Избор на ефективни формати за съхранение на данни
  - Преминаване от CSV и XLSX към по-ефективни формати (Apache Parquet, ORC) за съхраняване на големи количества данни
  - Внедряване на система за създаване на версии на данните за проследяване на промените
  - Използване на метаданни за описание на структурата и произхода на информацията
- Създаване на единна архитектура на фирмени данни
  - Сравняване на различни архитектури за съхранение: RDBMS, DWH и Data Lake. Изберете тази, която най-добре отговаря на нуждите ви от мащабируемост, интеграция на източници и аналитична обработка.
  - Проектирайте карта на процеса за извлечане, зареждане и трансформиране на данни (ETL) от различни източници за вашите задачи. Използвайте инструменти за визуализация, като Miro, Lucidchart или Draw.io, за да визуализирате ключови стъпки и точки на интеграция
- Прилагане на практики за управление на данните и минимализъм на данните
  - Следвайте подхода на минимализма на данните - съхранявайте и обработвайте само това, което е наистина ценно.
  - Прилагане на принципите за управление на данните - определяне на отговорността за данните, осигуряване на качество и прозрачност
  - Научете повече за политиките за управление на данни и концепциите на DataOps, VectorOps
  - Определяне на критерии за качество на данните и процедури за валидиране на данни в рамките на DataOps

Добре организираното съхранение на данни създава основата за централизиране на аналитичните процеси в компанията. Преходът от хаотично натрупване на файлове към структурирани хранилища позволява превръщането на информацията в стратегически актив, който помага за вземането на информирани решения и подобрява ефективността на бизнес процесите.

След като процесите на събиране, трансформиране, анализ и структурирано съхранение на данни ще бъдат автоматизирани и стандартизираны, следващият етап на цифровата трансформация е пълноценната работа с големи данни.





## **IX ЧАСТ**

# **ГОЛЕМИ ДАННИ, МАШИННО ОБУЧЕНИЕ И ПРОГНОЗИ**

Деветата част е посветена на големите данни, машинното обучение и прогнозния анализ в строителната индустрия. В нея се разглежда преходът от интуитивно вземане на решения към обективни анализи, базирани на исторически данни. Използват се практически примери, за да се демонстрира анализът на големи данни в строителството - от анализа на набора от данни за разрешителни за строеж в Сан Франциско до обработката на CAD - проекти с милиони елементи. Особено внимание е отделено на методите за машинно обучение за прогнозиране на разходите и сроковете за изпълнение на строителни проекти, като подробно са разгледани алгоритмите за линейна регресия и k-най-близки съседи. Показано е как структурираните данни се превръщат в основа за прогнозни модели за оценка на рисковете, оптимизиране на ресурсите и подобряване на ефективността на управлението на проекти. В частта се дават и препоръки за това как да се избират представителни извадки от данни и се обяснява защо за ефективните анализи не винаги са необходими големи набори от данни.

## ГЛАВА 9.1.

# ГОЛЕМИ ДАННИ И ТЕХНИЯТ АНАЛИЗ

### Големи данни в строителството: от интуиция към предвидимост

Терминът "големи данни" няма строго определение. Първоначално понятието се появява, когато обемът на информацията започва да надхвърля възможностите на традиционните методи за нейната обработка. Днес обемът и сложността на данните в много отрасли, включително в строителството, са се увеличили толкова много, че не се побират в локалната памет на компютрите и изискват използването на нови технологии за обработката им.

Същността на работата с големи данни е не само съхранението и обработката, но и възможностите за прогнозиране. В строителния бранш Големите данни отварят пътя от интуитивните решения, основани на субективно тълкуване на таблици и визуализации (както беше разгледано по-рано), към информирани прогнози, подкрепени от реални наблюдения и статистически данни.

Противно на общоприетото схващане, целта на работата с големи обеми данни не е "да накараме машината да мисли като човек", а да прилагаме математически модели и алгоритми за анализ на огромни количества данни, за да идентифицираме модели, да предсказваме събития и да оптимизираме процеси.

**Големите данни** не са студен свят на алгоритми, лишен от човешко влияние. Напротив, големите данни работят в комбинация с нашите инстинкти, грешки и творчество. Именно несъвършенството на човешкото мислене ни позволява да намираме нестандартни решения и да правим пробиви.

С развитието на цифровите технологии строителният бранш започна активно да използва техниките за обработка на данни, които идват от ИТ сектора. Благодарение на инструменти като Pandas и Apache Parquet, могат да се комбинират структурирани и неструктурни данни, като се опростява достъпът до информация и се намаляват загубите при анализ, а големите масиви от данни от документи или CAD проекти (фиг. 9.2-10 - фиг. 9.2-12) позволяват да се събират, анализират и прогнозират данни на всички етапи от жизнения цикъл на проекта.

Големите данни оказват трансформиращо въздействие върху строителния бранш, оказвайки му потенциално влияние по различни начини. Прилагането на технологиите за Big Data дава резултати в редица ключови области, включително например в следните:

- **Анализ на инвестиционния потенциал** - прогнозиране на рентабилността и сроковете на откупуване на проектите въз основа на данни от предишни обекти.
- **Предсказваща поддръжка** - идентифициране на вероятни повреди на оборудването, преди те да са настъпили, което намалява времето за престой.
- **Оптимизация на веригата за доставки** - прогнозиране на прекъсванията и подобряване на ефективността на логистиката.
- **Анализи на енергийната ефективност** - подпомагане на проектирането на сгради с ниска

консумация на енергия.

- **Мониторинг на безопасността** - използване на сензори и носими устройства за наблюдение на условията на обекта.
- **Контрол на качеството** - наблюдение в реално време на съответствието със стандартите на процеса.
- **Управление на работната сила** - анализ на изпълнението и прогнозиране на потребностите от персонал.

Трудно е да се намери област в строителството, в която анализът на данни и прогнозите да не са търсени. Основното предимство на алгоритмите за прогнозиране е способността им да се само-обучават и постоянно да се подобряват с натрупването на данни.

В близко бъдеще изкуственият интелект не само ще помага на строителите, но и ще взема ключови решения - от процесите на проектиране до въпросите, свързани с експлоатацията на сградите.

Повече за това как се генерират прогнози и как се използват моделите за обучение ще бъде разгледано в следващата част на книгата - "Машинно обучение и прогнози".

Преходът към пълноценна работа с големи данни изисква промяна в самия подход към анализа. Докато класическите системи, които разглеждахме досега, се фокусираха върху причинно-следствените връзки, анализът на големи данни измества фокуса си върху търсенето на статистически модели и корелации, които ни позволяват да идентифицираме скрити връзки и да прогнозираме поведението на обектите дори без пълно разбиране на всички фактори.

## **Въпросът за осъществимостта на големите данни: корелация, статистика и извадка от данни**

Традиционно конструирането се основава на субективни хипотези и личен опит. Инженерите предполагаха - с определена степен на вероятност - как ще се държи материалът, какви натоварвания ще издържи конструкцията и колко дълго ще продължи проектът. Тези предположения се прове-ряваха на практика, често с цената на време, ресурси и бъдещи рискове.

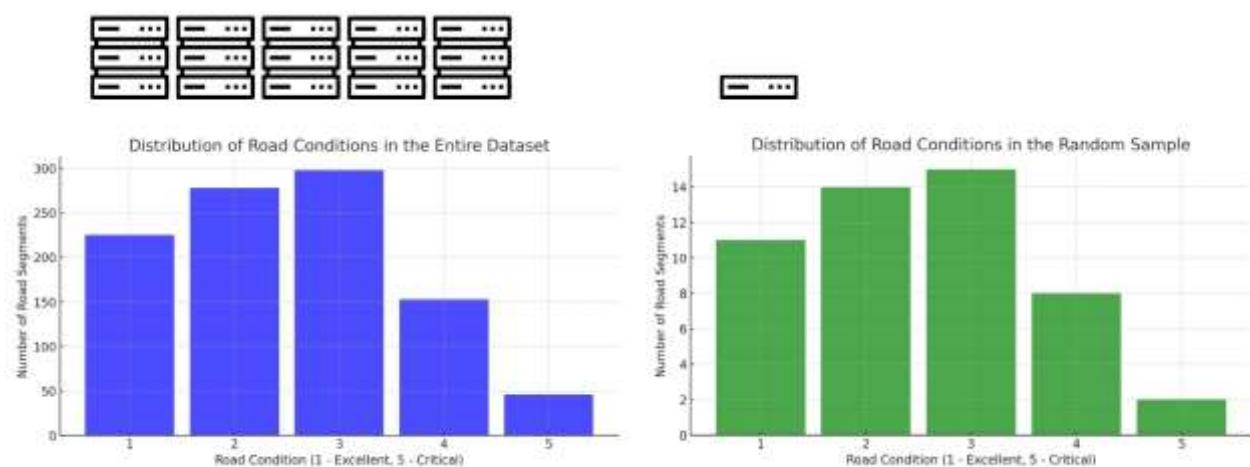
С навлизането на големите масиви от данни подходът се променя драстично: решението вече не се вземат въз основа на интуитивни догадки, а в резултат на анализ на големи масиви от данни. Строителството постепенно престава да бъде изкуство на интуицията и се превръща в точна наука за прогнозиране.

Преходът към идеята за използване на големи обеми от данни неизбежно повдига важен въпрос: доколко критичен е обемът на данните и колко информация е наистина необходима за надежден прогнозен анализ? Широко разпространеното убеждение, че "колкото повече данни, толкова по-голяма точност", не винаги се оказва статистически валидно на практика.

Още през 1934 г. статистикът Йежи Нойман доказва [145], че ключът към точността на статистическите изводи се крие не толкова в количеството данни, колкото в тяхната представителност и случайността на извадката.

Това важи особено за строителната индустрия, където се събират големи масиви от данни с помощта на IoT -сензори, скенери, камери за наблюдение, дронове и дори многоформатни CAD -модели, което увеличава риска от слепи петна, отклонения и изкривявания на данните.

Нека разгледаме пример за мониторинг на състоянието на пътната настилка. Пълният набор от данни за всички пътни участъци може да заема X GB и обработката му да отнеме около един ден. В същото време случайна извадка, включваща само всеки 50-ти пътен участък, ще отнеме само  $X/50$  GB и ще бъде обработена за половин час, като същевременно ще осигури подобна точност на оценките за определени изчисления (фигура 9.1-1).



**Фигура 9.1-1 Хистограми за състоянието на настилката: пълният набор от данни и случайната извадка показват идентични резултати.**

Следователно ключът към успешния анализ на данните често може да не е количеството данни, а представителността на извадката и качеството на използваните методи за обработка. Преминаването към случаини извадки и по-селективен подход изисква промяна в мисленето в строителната индустрия. В миналото компаниите са следвали логиката "колкото повече данни, толкова по-добре", вярвайки, че обхващането на всички възможни показатели ще увеличи точността.

Този подход напомня за една популярна заблуда от управлението на проекти: "Колкото повече специалисти привлека, толкова по-ефективна ще бъде работата". Въпреки това, както и при човешките ресурси, качеството и инструментите са по-важни от количеството. Без да се отчитат взаимовръзките (корелациите) между данните или участниците в проекта, увеличаването на обема може да доведе само до шум, изкривяване, дублиране и ненужни загуби.

В крайна сметка често се оказва, че е много по-продуктивно да се разполага с по-малък, но качествено подгответен набор от данни, който да може да даде стабилни и разумни прогнози, отколкото да се разчита на огромна, но хаотична информация, съдържаща много противоречиви сигнали.

Прекомерният обем данни не само не гарантира по-голяма точност, но може да доведе и до изкривени заключения поради наличието на шум, излишни характеристики, скрити корелации и нерелевантна информация. При такива обстоятелства се увеличава рисъкът от прекомерно приспособяване на моделите и се намалява надеждността на аналитичните резултати.

В строителството основно предизвикателство при работата с големи обеми данни е определянето на оптималното количество и качество на данните. Например при наблюдение на състоянието на бетонни конструкции използването на хиляди сензори и събирането на информация всяка минута може да претовари системата за съхранение и анализ. Ако обаче се извърши корелационен анализ и се изберат 10% от най-информативните сензори, може да се получи почти идентична точност на прогнозите, като се изразходват многократно, понякога десетки и стотици пъти, по-малко ресурси.

Използването на по-малка подгрупа от данни намалява както обема на необходимото пространство за съхранение, така и времето за обработка, което значително намалява разходите за съхранение и анализ на данни и често прави случайната извадка идеално решение за прогнозен анализ, особено при големи инфраструктурни проекти или при работа в реално време. В крайна сметка ефективността на строителните процеси не се определя от количеството събрани данни, а от качеството на техния анализ. Без критичен подход и внимателен анализ данните могат да доведат до погрешни заключения.

След определено количество данни всяка нова единица информация дава все по-малко полезни резултати. Вместо безкрайно да събирате информация, важно е да се съсредоточите върху нейната представителност и методите за анализ (фигура 9.2-2).

Този феномен е добре описан от Алън Уолис [146], който илюстрира използването на статистически методи с примера на тестването на два алтернативни дизайна на снаряди на военноморските сили на САЩ.

Военноморските сили изprobваха два алтернативни дизайна на снаряди (A и B), като проведоха серия от сдвоени изстрели. Във всеки рунд A получава 1 или 0 в зависимост от това дали представянето му е по-добро или по-лошо от това на B, и обратно. Стандартният статистически подход включва провеждане на фиксиран брой опити (например 1000) и определяне на победителя въз основа на процентно разпределение (например, ако A получава 1 в повече от 53% от случаите, той се счита за най-добър). Когато Алън Уолис обсъжда такъв проблем с капитана от (военноморските) сили Гарет Л. Шуйлер, капитанът възразява, че такъв тест, ако цитираме историята на Алън, може да е безполезен. Ако на място беше мъдър и опитен офицер по боеприпасите като Шуйлер, той щеше да види след първите няколкостотин [изстрела], че експериментът не е необходимо да се прекратява, било защото новият метод е явно по-лош, било защото е явно по-добър от това, на което се е надявал [146].

- Статистическа изследователска група на правителството на САЩ в Колумбийския университет, период на Втората световна война

Този принцип се използва широко в различни индустрии. В медицината например клиничните изпитвания на нови лекарства се провеждат върху случаини извадки от пациенти, което позволява да се получат статистически значими резултати, без да е необходимо лекарството да се тества върху цялата популация от хора, живеещи на планетата. В икономиката и социологията се провеждат представителни проучвания, за да се отрази мнението на обществото, без да е необходимо да се интервюират всички жители на страната.

Точно както правителствата и изследователските организации проучват малки групи от населението, за да разберат общите социални тенденции, компаниите в строителния бранш могат да използват случаини извадки от данни, за да извършват ефективен мониторинг и да създават прогнози за управлението на проекти (Фигура 9.1-1).

*Големите данни могат да променят подхода към социалните науки, но няма да заменят статистическия здрав разум [147].*

- Thomas Landsall-Welfair, "Прогнозиране на текущото настроение на нацията", Significance v. 9(4), 2012 г.

От гледна точка на пестенето на ресурси, когато се събират данни за бъдещи прогнози и вземане на решения, е важно да се отговори на въпроса: има ли смисъл да се изразходват значителни ресурси за събиране и обработка на огромни набори от данни, когато може да се използва много по-малък и по-евтин набор от тестови данни, който може да се увеличава постепенно? Ефективността на случаината извадка показва, че компаниите могат да намалят разходите за събиране и обучение на модели с десетки или дори хиляди пъти, като изберат методи за събиране на данни, които не изискват цялостно покритие, но все пак осигуряват достатъчна точност и представителност. Този подход позволява дори на малки компании да постигат резултати, равни на тези на големите корпорации, като използват значително по-малко ресурси и обеми от данни, което е важно за компаниите, които искат да оптимизират разходите и да ускорят вземането на информирани решения с помощта на малки ресурси. В следващите глави разгледайте примери за анализи и прогнозни анализи, базирани на публични масиви от данни, използвайки инструменти за големи данни.

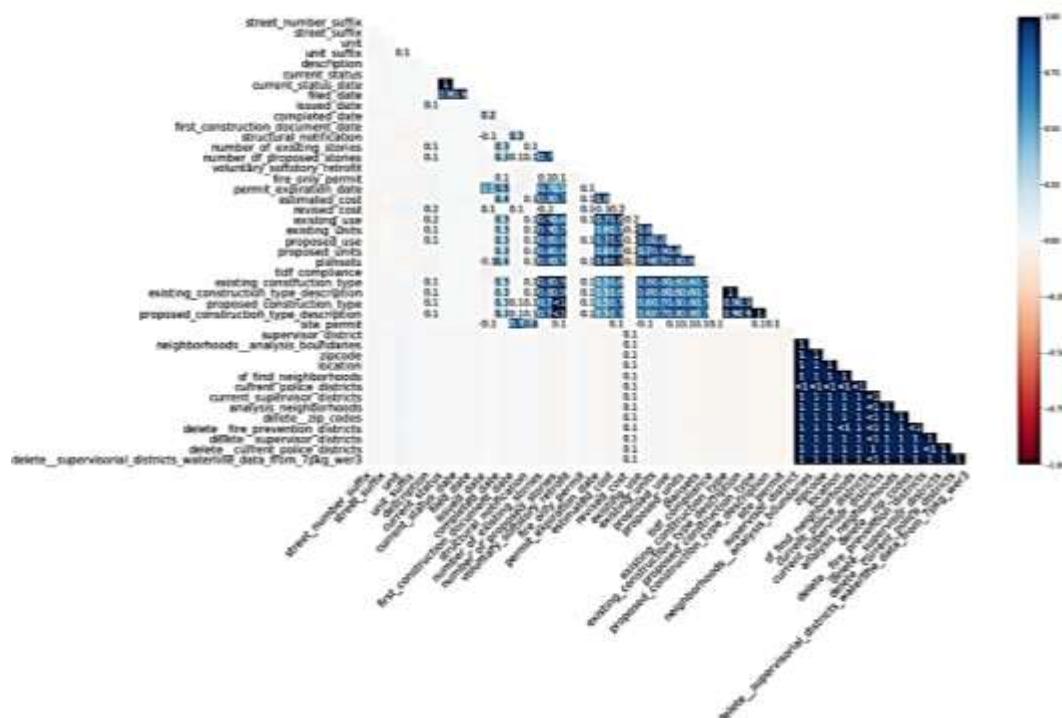
## **Големи данни: анализ на данни от набора от данни за милиони разрешителни за строеж в Сан Франциско**

Работата с отворени масиви от данни предоставя уникална възможност да се приложат на практика принципите, разгледани в предишните глави: разумен подбор на характеристики, представителна извадка, визуализация и критичен анализ. В тази глава ще проучим как сложни явления като строителната дейност в голям град могат да бъдат изследвани с помощта на отворени данни - по-специално над един милион записи на разрешителни за строеж в Сан Франциско.

Публично достъпни данни за над един милион разрешителни за строеж (фигура 9.1-2) (записи в два набора от данни във формат CSV) от "San Francisco Department of Buildings" [148] ни позволяват да използваме необработената CSV -таблица, за да анализираме не само строителната дейност в града, но и да анализираме критично последните тенденции и историята на строителната индустрия на Сан Франциско през последните 40 години, от 1980 до 2019 г.

Примерите за код, използвани за създаване на визуализациите на наборите от данни (фигури 9.1-3- фигури 9.1-8), както и визулните графики с код, обяснения и коментари, могат да бъдат намерени в платформата Kaggle чрез търсене на "San Francisco. Строителен сектор 1980-2019 г." [149].

Фиг. 9.1-2 Наборите от данни съдържат информация за издадени разрешителни за строеж с различни атрибути на обектите.



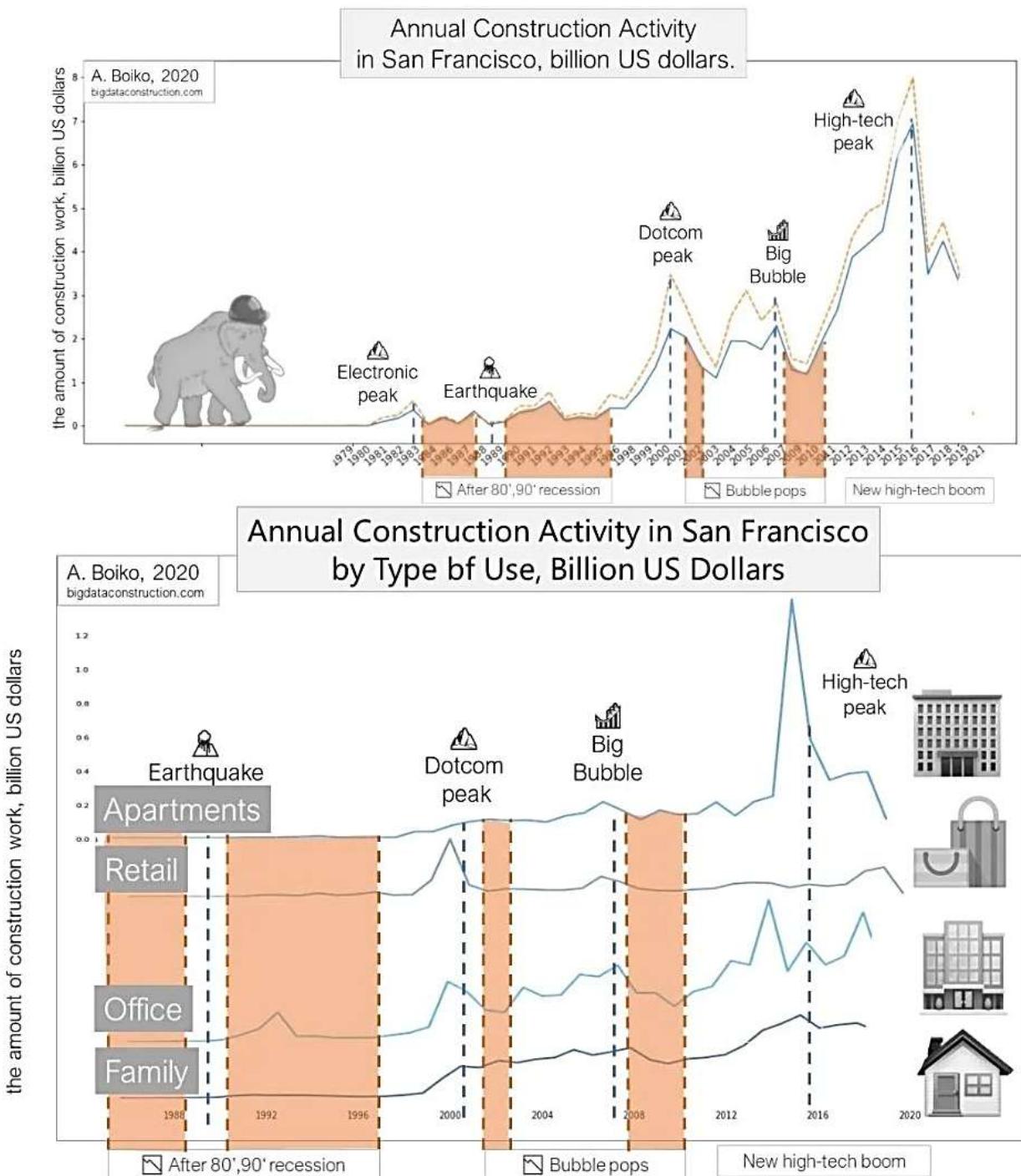
Фиг. 9.1-3 Топлинна карта (Pandas и Seaborn), която визуализира всички атрибути на набор от данни и помага да се определят връзките между двойки атрибути.

От таблицата, предоставена от Департамента по строителство на Сан Франциско (фигура 9.1-2), не

се виждат никакви тенденции или заключения. Сухите числа в табличен вид не са основа за вземане на решения. За да станат данните визуално разбираеми, както е разгледано подробно в главите за визуализация на данните, те трябва да се визуализират с помощта на различните библиотеки, разгледани в седма част на книгата на тема "ETL и визуализиране на резултатите като графики".

Анализирайки данните, използвайки библиотеките за визуализация на Pandas DataFrame и Python стойността на 1 137 695 разрешителни [148], можем да заключим, че строителната дейност в Сан Франциско е тясно свързана с икономическите цикли, особено в процъфтяващата технологична индустрия в Силициевата долина (фигура 9.1-4).

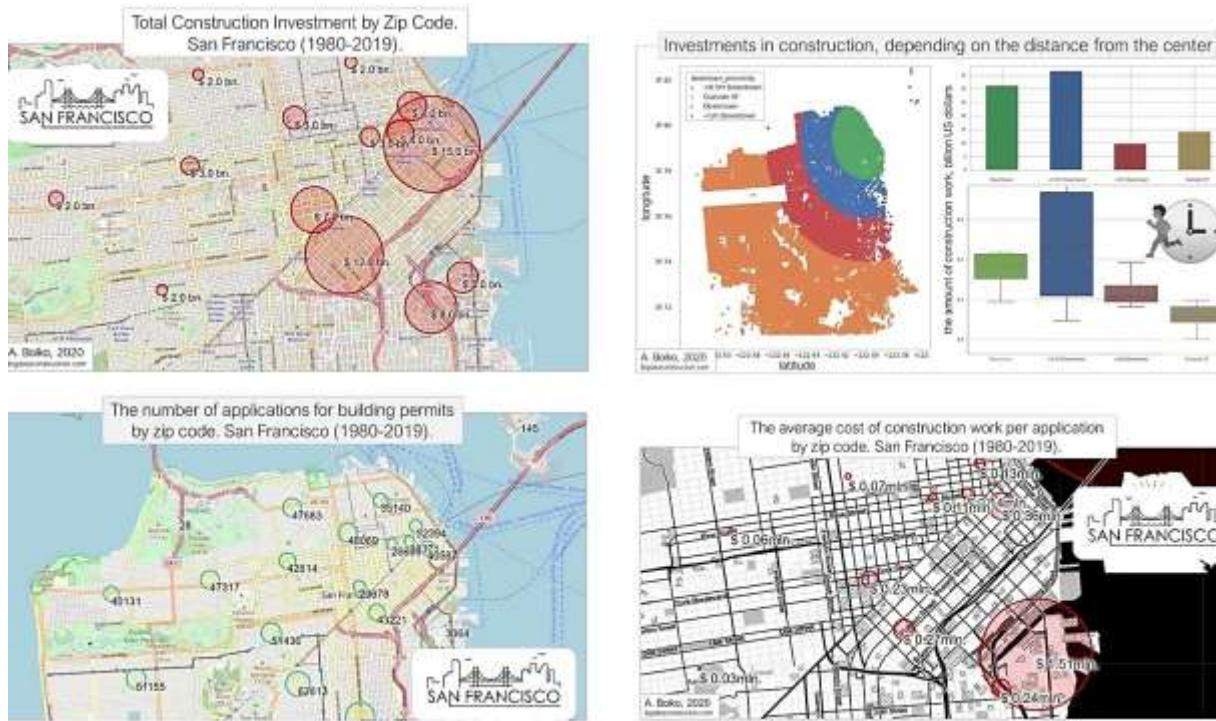
Икономическите бумове и спадове оказват значително влияние върху броя и стойността на строителните проекти. Например първият пик в строителната дейност съвпада с бума на електрониката в средата на 80-те години на ХХ век (използвани са Pandas и Matplotlib), а последващите пикове и спадове са свързани с дотком балона и технологичния бум от последните години.



Фиг. 9.1-4 В сектора на недвижимите имоти в Сан Франциско инвестициите корелират с технологичното развитие на Силициевата долина.

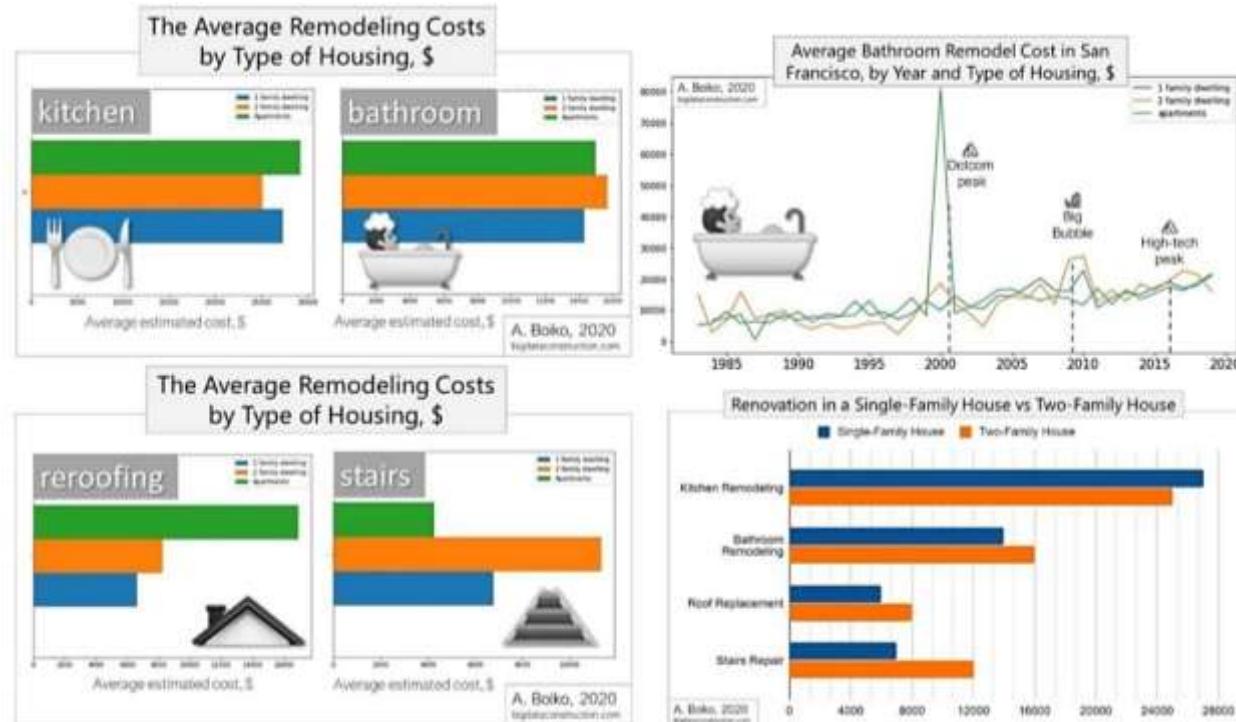
Анализът на данните показва, че в Сан Франциско по-голямата част от 91,5 млрд. долара, инвестиирани в строителство и преустройство през последното десетилетие - почти 75% - са концентрирани в центъра на града (фигура 9.1-5 - използвани са Pandas и библиотеката за визуализация Folium) и в радиус от 2 км от центъра на града, което отразява по-високата плътност на инвестициите в тези централни зони.

Средните разходи за разрешителни за строеж варират значително в зависимост от квартала, като в центъра на града те са три пъти по-скъпи от тези извън него поради по-високите разходи за земя, труд, материали и строгите строителни разпоредби, изискващи по-скъпи материали за подобряване на енергийната ефективност.



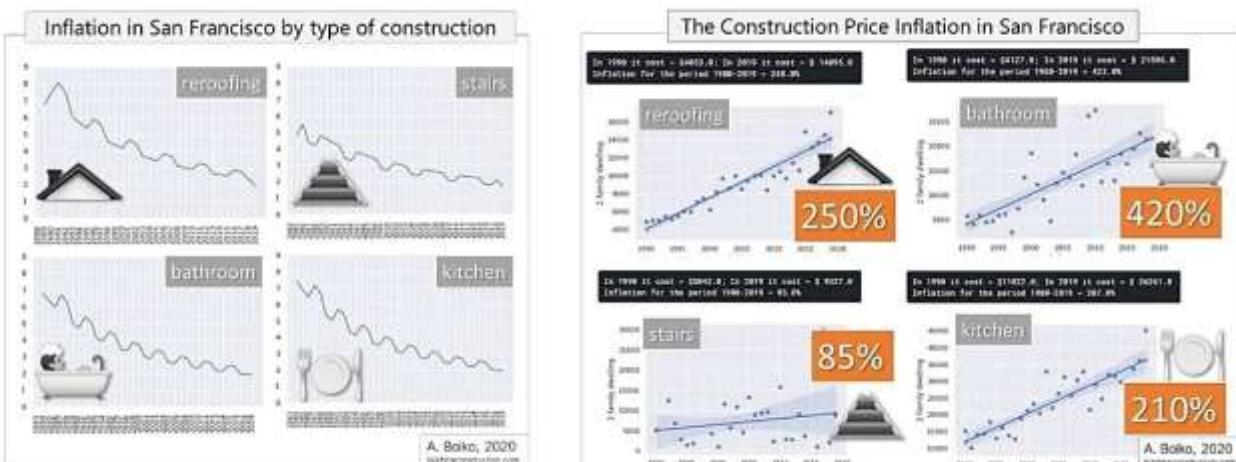
Фигура 9.1-5 В Сан Франциско 75% от строителните инвестиции (91,5 млрд. долара) са съсредоточени в центъра на града.

Наборът от данни позволява също така да се изчислят средните цени на ремонтите не само по видове къщи, но и по квартали на града и отделни адреси (пощенски кодове). В Сан Франциско динамиката на разходите за ремонт на жилища показва различни тенденции за различните видове ремонти и жилища (фигура 9.1-6 - използвани са Pandas и Matplotlib). Ремонтът на кухнята е забележимо по-скъп от ремонта на банята: средният ремонт на кухнята в еднофамилна къща струва около 28 000 USD в сравнение с 25 000 USD в двуфамилна къща.



Фигура 9.1-6 В СФ ремонтът на кухнята струва почти два пъти по-скъпо от ремонта на банята, а собствениците на жилища трябва да заделят \$ 350 всеки месец в продължение на 15 години, за да покрият разходите за основен ремонт на жилището.

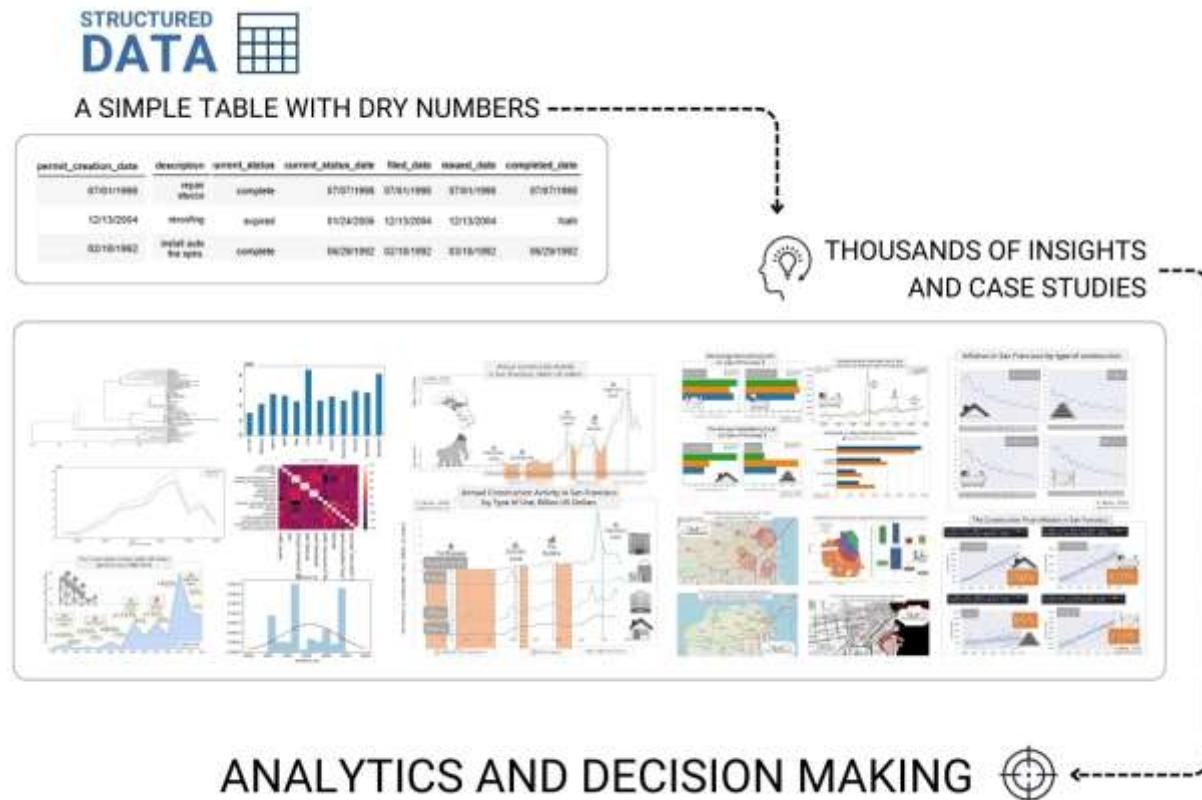
Инфлацията на разходите за строителство в Сан Франциско през годините може да се проследи чрез анализ на данните, групирани по видове жилища и години (фигура 9.1-7 - използвана от Pandas и Seaborn), която показва постоянно нарастващо на средните разходи за ремонт от 1990 г. насам и разкрива краткосрочни тригодишни цикли в разходите за ремонт на многофамилни сгради.



Фигура 9.1-7 От 1980 г. до 2019 г. разходите за ремонт на банята в СФ са се увеличили пет пъти, докато ремонтът на покрива и кухнята е поскъпнал три пъти, а ремонтът на стълбището е поскъпнал само с 85%.

Проучването на публични данни от Строителния отдел на Сан Франциско (фигура 9.1-3) разкрива, че разходите за строителство в града са изключително променливи и често непредвидими, повлияни от различни фактори. Тези фактори включват икономическия растеж, технологичните иновации и уникалните изисквания на различните видове жилища.

В миналото подобни анализи изискваха задълбочени познания в областта на програмирането и анализа. С появата на инструментите LLM- обаче процесът стана достъпен и разбираем за широк кръг специалисти в строителния бранш - от инженерите в проектантските отдели до висшия менеджмънт.



**Фиг. 9.1-8** Преминаването към визуално разбирами данни дава възможност за автоматизирано вземане на решения чрез разпознаване на скрити модели.

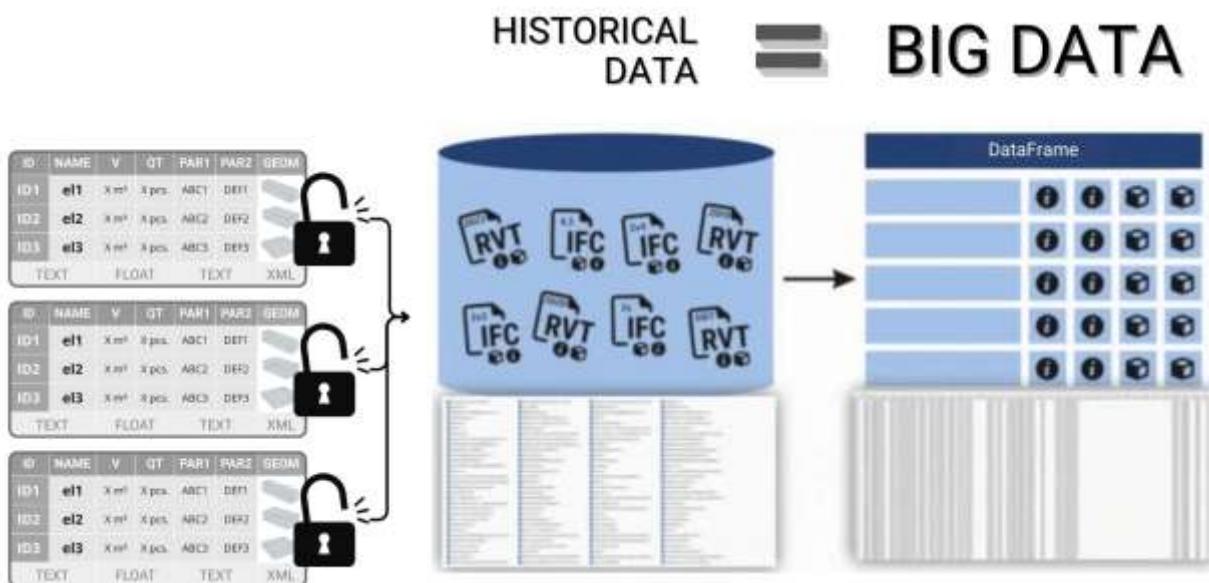
Точно както анализирахме данните от табличния набор от данни на "San Francisco Building Authority", можем да визуализираме и анализираме всеки набор от данни - от изображения и документи до данни от IoT, или данни от производни бази данни на CAD.

### Пример за големи данни, базирани на CAD данни (BIM)

В следващия пример ще анализираме голям набор от данни, като използваме данни от различни CAD инструменти (BIM). За събирането и създаването на голямата съвкупност от данни беше използван специализиран автоматичен уебпотърсач (скрипт), конфигуриран за автоматично търсене и събиране на проектни файлове от уебсайтове, предлагащи безплатни архитектурни модели

във форматите RVT и IFC. В рамките на няколко дни обхождащата програма успешно намери и изтегли 4596 IFC файла, 6471 RVT файла и 156 024 DWG файла[149].

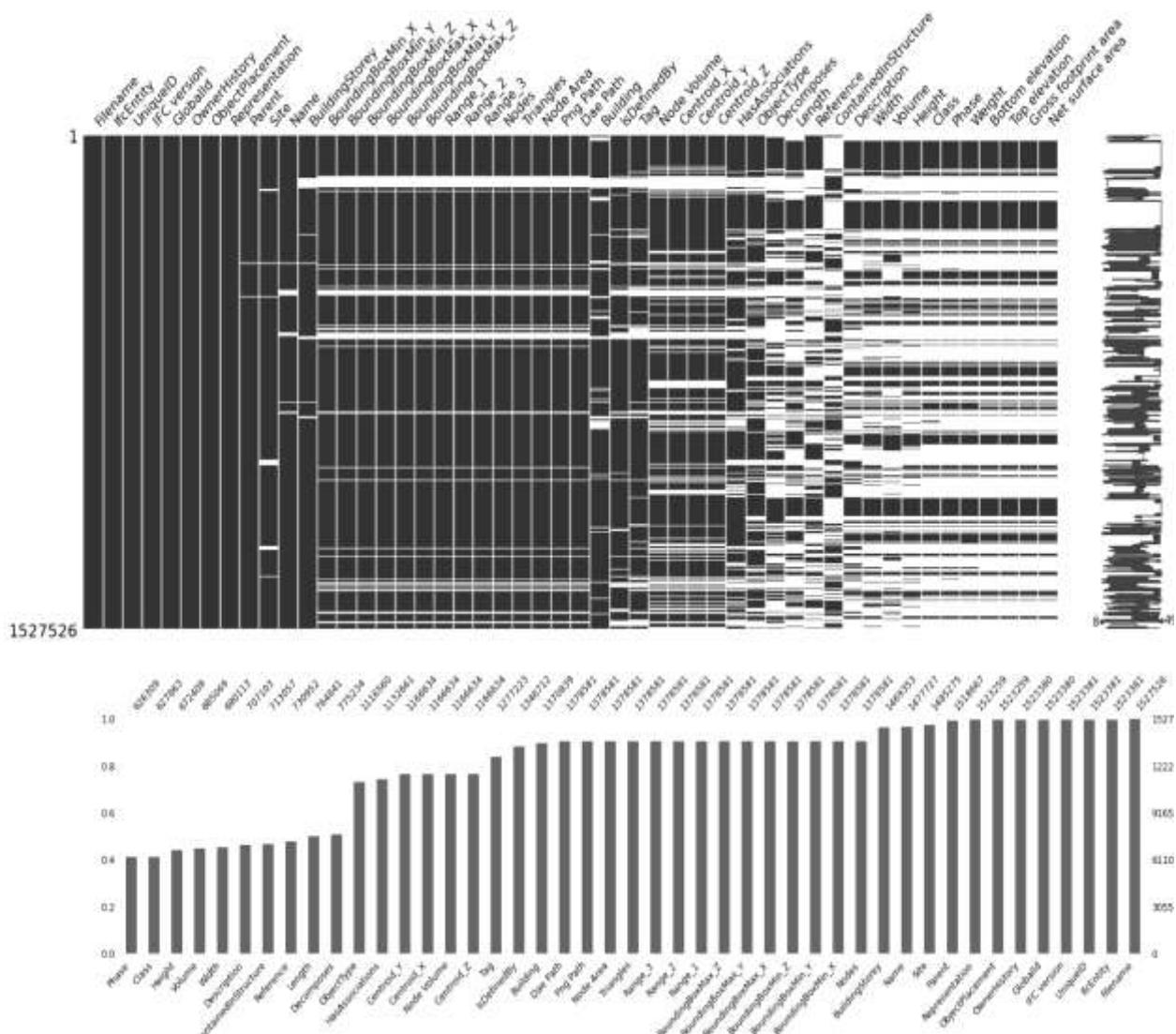
След като бяха събрани проекти във формати RVT и IFC от различни версии и бяха преобразувани в структуриран CSV формат с помощта на бесплатните SDK за обратно инженерство, почти 10 хиляди проекта RVT и IFC бяха събрани в един голям файл с таблица Apache Parquet и качени в Pandas DataFrame за анализ (Фигура 9.1-9).



**Фигура 9.1-9 Структурирани данни за проекти ви позволява да комбинирате произволен брой проекти в една двуизмерна таблица.**

Данните от тази широкомащабна колекция съдържат следната информация: наборът от файлове IFC съдържа около 4 miliona същности (редове) и 24 962 атрибути (колони), а наборът от файлове RVT, състоящ се от около 6 miliona същности (редове), съдържа 27 025 различни атрибути (колони).

Тези информационни масиви (фигура 9.1-10) обхващат милиони елементи, за всеки от които допълнително са получени координатите на геометрията Bounding Box (правоъгълник, определящ границите на обект в проекта) и са добавени в общ табличен формат - координатите на геометрията Bounding Box (правоъгълник, определящ границите на обект в проекта) и са създадени изображения на всеки елемент във формат PNG и геометрията в отворения формат XML - DAE (Collada).



Фигура 9.1-10 Subset от 1,5 miliona елемента и визуализация (библиотека missingno) на заетостта на първите 100 атрибути като хистограма.

По този начин получихме цялата информация за десетки милиони елементи от 4596 проекта IFC и 6471 проекта RVT, където всички атрибути-свойства на всички елементи на същността и тяхната геометрия (Bounding Box) бяха преведени в структурирана форма на една таблица (DataFrame) (фиг. 9.1-10 - данните за популациите на dataframe изглеждат като хистограми).

Хистограмите (фиг. 9.1-10, фиг. 9.2-6, фиг. 9.2-7), начертани по време на процеса на анализ, позволяват бърза оценка на гъстотата на данните и честотата на срещане на стойностите в колоните. Това дава първа представа за разпределението на признаките, наличието на отклонения и потенциалната полезност на отделните атрибути при анализа и изграждането на модели за машинно обучение.

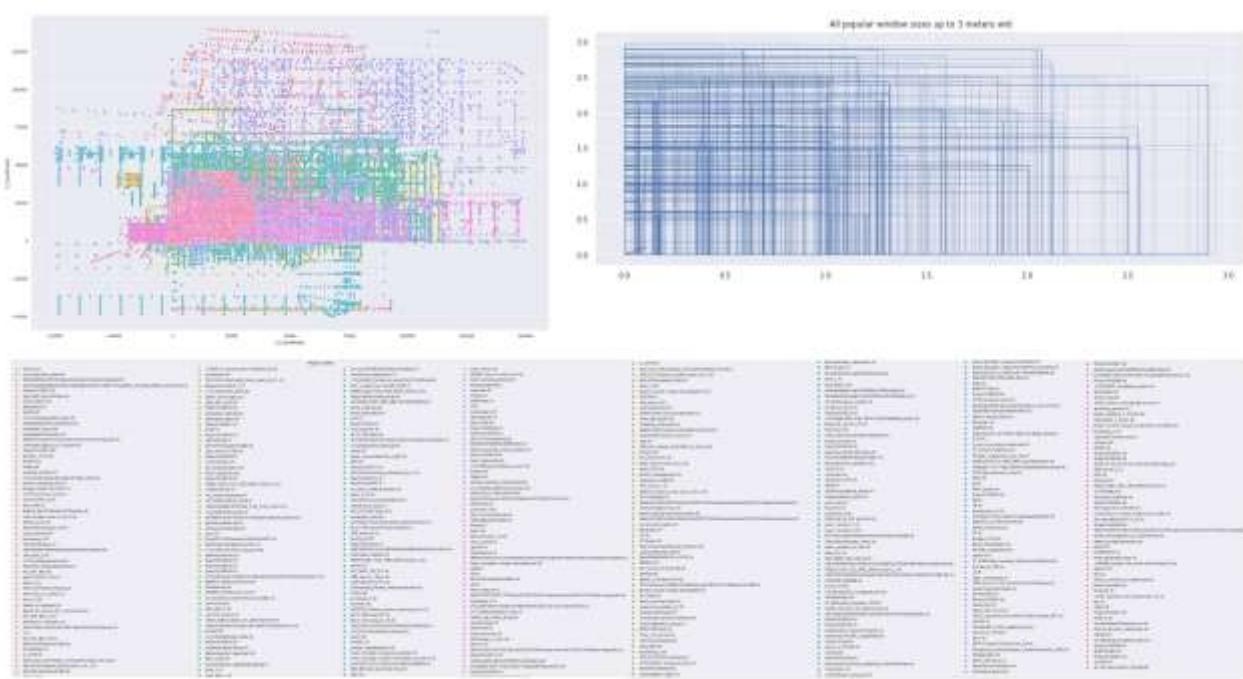
Един пример за практическо използване на този набор от данни (фиг. 9.1-10) е проектът "5000 IFC и RVT ". [149], достъпен в платформата Kaggle. Той представя Jupyter Notebook с цялостно решение

за Pipeline: от предварителна обработка и анализ на данните до визуализация на резултатите с помощта на библиотеки на Python - pandas, matplotlib, seaborn, folium и други (фиг. 9.1-11).



**Фигура 9.1-11 Примери за анализ на данни от CAD формати (BIM) с помощта на библиотеки за визуализация на Python и библиотеката pandas.**

Въз основа на метаданните е възможно да се определи в кои градове са разработени определени проекти и това да се изобрази на карта (напр. с помощта на библиотеката folium). Освен това времевите печати в данните позволяват да се изследват моделите във времето, когато файловете са били запазени или редактирани: по дни от седмицата, часове от деня и месеци.



**Фиг. 9.1-12 Визуализация на геометричното положение на всички колони и размерите на всички прозорци до 3 метра в проектите от списъка в долната част на диаграмата.**

Геометричните параметри под формата на Bounding Box, извлечени от моделите, също се поддават

на съвкупен анализ. Например на фигура 9.1-12 са показани две графики: лявата показва разпределението на разстоянията между колоните за всички проекти спрямо нулевата точка, а дясната показва размерите на всички прозорци с височина до 3 метра в извадка от десетки хиляди прозоречни елементи (след групиране на целия набор от данни по параметъра "Категория" със стойност "OST\_Windows", "IfcWindows").

Кодът за анализ на тръбопроводите за този пример и самият набор от данни са достъпни на уебсайта на Kaggle под заглавието "5000 IFC и RVT | DataDrivenConstruction.io projects" [149]. Този завършен Pipeline заедно с набора от данни може да бъде копиран и стартиран бесплатно онлайн безплатно на Kaggle или онлайн в някоя от популярните IDE: PyCharm, Visual Studio Code (VS Code), Jupyter Notebook, Spyder, Atom, Sublime Text, Eclipse с PyDev плъгин, Thonny, Wing IDE, IntelliJ IDEA с Python плъгин, JupyterLab или популярните онлайн инструменти Kaggle.com, Google Collab, Microsoft Azure Notebooks, Amazon SageMaker.

Аналитичните прозрения, получени от обработката и изучаването на огромни количества структурирани данни, ще играят решаваща роля в процесите на вземане на решения в строителната индустрия.

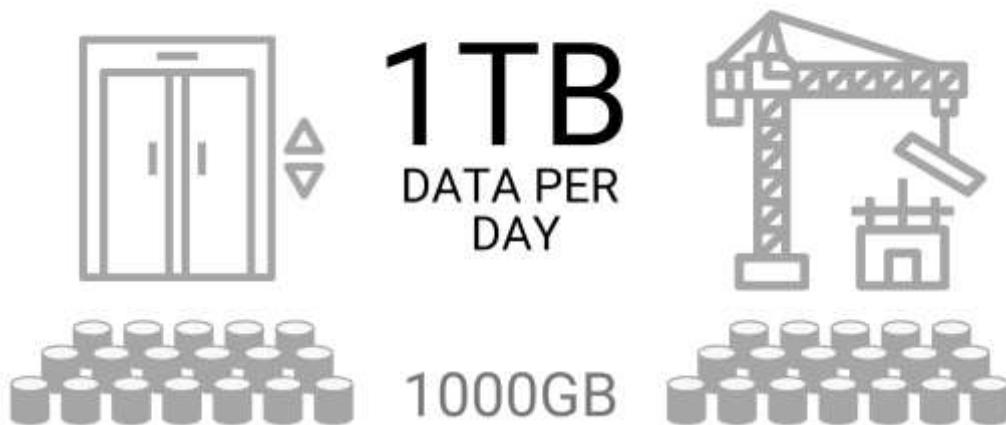
С помощта на този вид информационен анализ, базиран на предишни проекти, специалистите могат ефективно да прогнозират например нуждите от материали и работна ръка и да оптимизират проектните решения преди началото на строителството

Но докато данните от проектирането или разрешителните за строеж са относително статична информация, която се променя сравнително бавно, самият строителен процес бързо се насища с разнообразни сензори и IoT -устройства: камери, автоматизирани системи за наблюдение, които предават данни в реално време - всичко това превръща строителната площадка в динамична цифрова среда, в която данните трябва да се анализират в реално време.

## IoT Интернет на нещата и интелигентни договори

IoT Интернет на нещата представлява нова вълна на цифрова трансформация, при която всяко устройство получава собствен IP адрес и става част от глобална мрежа. IoT е концепция, която включва свързване на физически обекти към интернет с цел събиране, обработка и предаване на данни. В строителството това означава възможност за контрол на строителните процеси в реално време, минимизиране на материалните отпадъци, прогнозиране на износването на оборудването и автоматизиране на процеса на вземане на решения.

Според статията на CFMA "Подготовка за бъдещето със свързаното строителство" [150], през следващото десетилетие строителната индустрия ще претърпи сериозна цифрова трансформация, която ще завърши с концепцията за Connected Construction - напълно интегрирана и автоматизирана строителна площадка.



Фиг. 9.1-13 Устройствата за данни на IoT или на строителната площадка могат да произвеждат и предават терабайти данни на ден.

Цифровата строителна площадка означава, че всички елементи на строителството - от планирането и логистиката до изпълнението на работата и контрола на качеството на строителната площадка с помощта на стационарни камери и квадрокоптери - ще бъдат интегрирани в една динамична цифрова екосистема. По-рано, в част 7 на тази книга, вече разглеждахме възможностите на Apache NiFi (фигура 7.4-5) - безплатен инструмент с отворен код, който позволява поточно предаване на данни в реално време - от събирането им от различни източници до прехвърлянето им към платформи за съхранение или анализ.

Данните за хода на строителството, разхода на материали, състоянието на оборудването и безопасността ще се предават в реално време към аналитичните системи (фиг. 9.1-13). Това позволява да се прогнозират потенциалните рискове, да се реагира своевременно на отклоненията и да се оптимизират процесите на обекта. Ключовите компоненти на една цифрова строителна площадка включват:

- IoT - сензори - проследяване на параметрите на околната среда, наблюдение на строителни машини и контрол на условията на труд.
- Цифрови близнаци - виртуални модели на сгради и инфраструктура за предвиждане на възможни отклонения и предотвратяване на грешки.
- Автоматизирани логистични системи - управление на веригата за доставки в реално време за намаляване на престоя и разходите.
- Работализирани строителни системи - използване на автономни машини за изпълнение на рутинни и опасни задачи.

Работализацията използване на IoT и концепцията за дигитална строителна площадка Connected Site (Construction) не само ще повишат ефективността и ще намалят разходите, но и ще поставят началото на нова ера на безопасност, устойчиво строителство и прогнозно управление на проекти.

Таговете RFID (Radio Frequency Identification) също са един от най-важните компоненти на IoT. Те се използват за идентифициране и проследяване на материали, машини и дори персонал на строителната площадка, като увеличават прозрачността и контрола на ресурсите на проекта.

RFID -технологията се използва за автоматично разпознаване на обекти с помощта на радиосигнали. Тя се състои от три основни елемента:

- RFID - етикети (пасивни или активни) - съдържат уникален идентификатор и се прикрепят към материали, инструменти или машини.
- Сканерите са устройства, които четат информация от етикетите и я предават на системата.
- Централизирана база данни - съхранява информация за местоположението, състоянието и движението на обектите.

Приложение на RFID в строителството:

- Автоматично отчитане на материалите - етикетите върху готовите бетонови изделия, арматурата или пакетите със сандвич панели позволяват контрол на наличностите и предотвратяват кражби.
- Контрол на работата на персонала - RFID - баджовете на служителите записват началния и крайния час на смяната, като осигуряват запис на работното време.
- Мониторинг на оборудването - RFID - система за проследяване на движението на оборудването, която предотвратява престой и подобрява ефективността на логистиката.

Този технологичен пакет се допълва от интелигентни договори, базирани на блокчейн, които автоматизират плащанията, контрола на доставките и спазването на договорите без необходимост от посредници, като намаляват риска от измами и забавяния.

Днес, при липсата на общ модел на данните, интелигентните договори са просто код, който участниците съгласуват. С подхода, ориентиран към данните, обаче е възможно да се създаде общ модел на параметрите на договора, да се кодира в блокчейн и да се автоматизира изпълнението на условията.

Например в система за управление на веригата за доставки интелигентен договор може да проследява доставката на пратка от сензори на IoT и етикети RFID и автоматично да превежда плащане, когато тя пристигне. По подобен начин на строителна площадка интелигентен договор може да регистрира завършването на даден етап от работата - например инсталацирането на арматурни заготовки или изливането на основи - въз основа на данни от безпилотни самолети или строителни сензори и автоматично да инициира следващото плащане към изпълнителя, без да са необходими ръчни проверки и хартиени сертификати.

Но въпреки новите технологии и усилията на международните организации за стандартизация, множеството конкуриращи се стандарти усложняват пейзажа на интернет на нещата.

Според проучване на Cisco, публикувано през 2017 г. [151], почти 60% от инициативите в областта на интернет на нещата (IoT) спират на етапа на доказване на концепцията, а само 26% от компаниите смятат, че техните проекти в областта на интернет на нещата са напълно успешни. Нещо повече, една трета от завършените проекти не постигат поставените цели и не се признават за успешни дори след изпълнението им.

Една от основните причини е липсата на оперативна съвместимост между платформите, които обработват данни от различни сензори. В резултат на това данните остават изолирани в рамките на

отделни решения. Алтернатива на този подход, както и в други подобни случаи (които сме разгледали в тази книга), е архитектура, изградена около самите данни като основен актив.

Сензорите на интернет на нещата играят ключова роля не само при наблюдението на техническото състояние на оборудването, но и при прогнозния анализ за намаляване на рисковете на строителната площадка и подобряване на цялостната ефективност на процесите чрез предвиждане на повреди и отклонения.

Данните, събрани от сензорите на IoT и етикетите RFID, могат да бъдат обработвани в реално време от алгоритми за машинно обучение, които могат да откриват аномалии и да предупреждават инженерите за потенциални неизправности предварително. Те могат да варират от микропукнатини в бетонни конструкции до нехарактерни паузи в работата на кулокрана, които показват технически неизправности или регуляторни нарушения. Освен това усъвършенстваните алгоритми за поведенчески анализ могат да улавят поведенчески модели, които могат да показват например физическа умора на персонала, като подобряват проактивното управление на безопасността и благосъстоянието на служителите на обекта.

В строителния бранш злополуките и неизправностите - независимо дали става въпрос за машини или хора - рядко се случват внезапно. Обикновено те са предшествани от малки отклонения, които остават незабелязани. Предсказващият анализ и машинното обучение позволяват да се откриват тези сигнали на ранен етап, дори преди да настъпят критични последици.

Докато документите, проектните файлове и данните от устройствата IoT и етикетите RFID формират цифровия отпечатък на строителните проекти, машинното обучение може да помогне за извличането на полезна информация от тях. С нарастването на обема на данните и демократизирането на достъпа до тях строителната индустрия получава нови възможности в областта на анализите, прогнозните анализи и приложенията на изкуствения интелект.

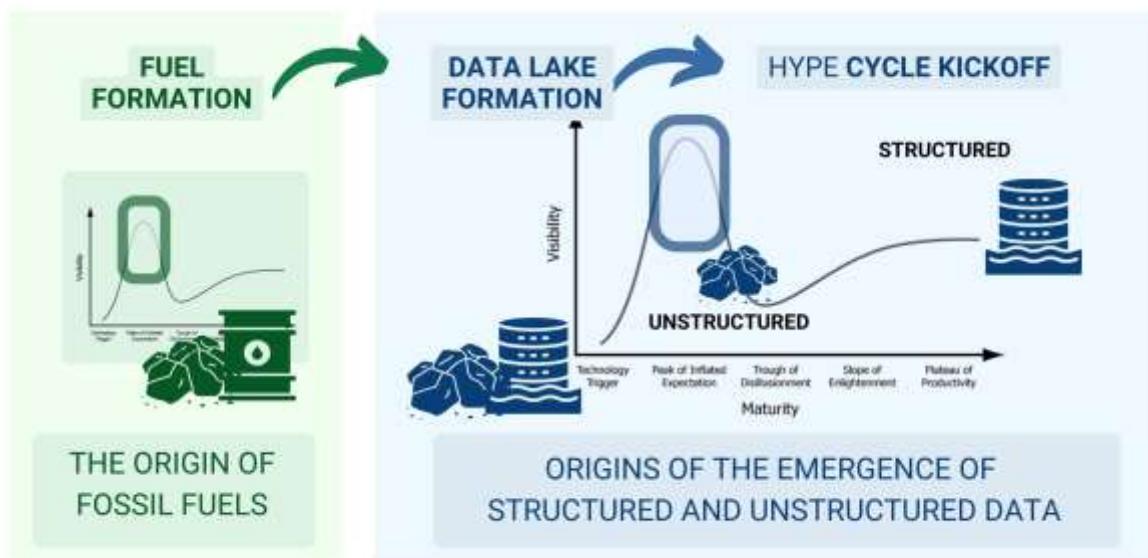


## ГЛАВА 9.2.

# МАШИННО ОБУЧЕНИЕ И ПРОГНОЗИ

### Машинното обучение и изкуственият интелект ще променят начина, по който създаваме

Базите данни на различните системи в строителния бизнес - с тяхната неизбежно разпадаща се и все по-сложна инфраструктура - се превръщат в хранителна среда за бъдещи решения. Фирмените сървъри, подобно на гората, са богати на биомаса от важна информация, често скрита под земята, в дебрите на папките и сървърите. Масивите от данни от различните системи, които се създават днес - след употреба, след падане на дъното на сървъра и след години на вкаменяване - ще подхранват машинното обучение и езиковите модели в бъдеще. Вътрешнофирмените чатове (напр. отделен екземпляр на локално конфигурираните ChatGPT, LLaMa, Mistral, DeepSeek) ще бъдат изградени на базата на тези вътрешни модели, използвани централизирано хранилище за бързо и удобно извлечение на информация и генериране на необходимите графики, информационни табла и документи.



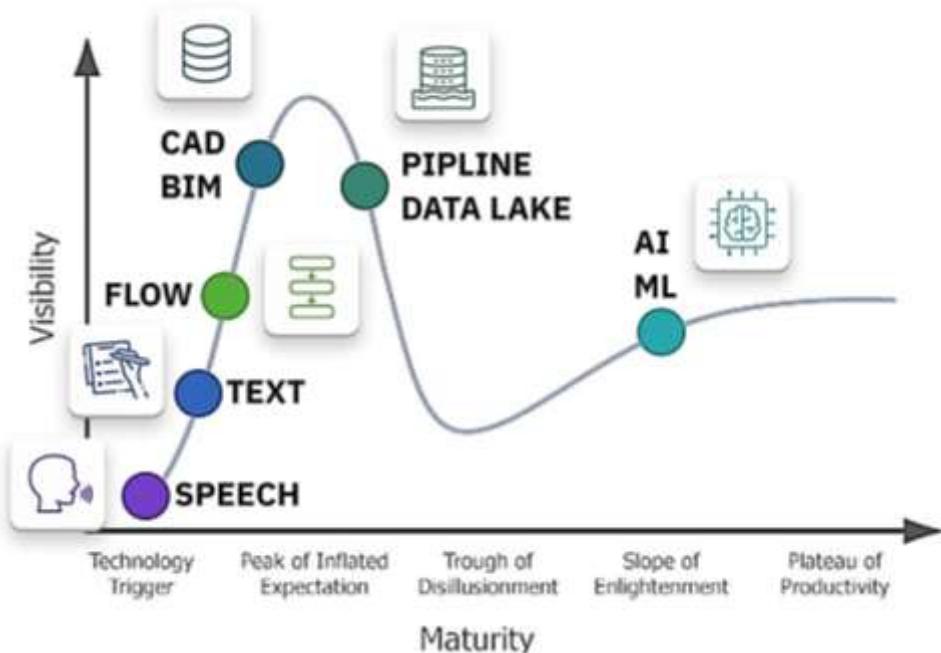
Фиг. 9.2-1 Точно както дърветата се превръщат във въглища, така и информацията се превръща в ценна бизнес енергия с течение на времето под натиска на времето и анализите.

Вкаменяването на растителната маса в комбинация с налягането и температурата създава хомогенна и уникално структурирана еднородна маса от дървeta от различни видове, живели по различно време - дървени въглища [152]. По същия начин информацията, записана на твърди дискове в различни формати и по различно време под натиска на аналитичните отели и температурата на управлението на качеството, в крайна сметка образува хомогенна структурирана маса от ценна информация (фиг. 9.2-1).

Тези слоеве (или по-често изолирани самородни парченца) информация се създават чрез старательно организиране на данните от опитни анализатори, които постепенно започват да извличат ценна информация от привидно отдавна несъществени данни.

В момента, в който тези зрели слоеве от данни вече не са просто "изгорени" в отчетите, а започват да циркулират в бизнес процесите, обогатявайки решенията и подобрявайки процесите, компанията става готова за следващата стъпка - преход към машинно обучение и изкуствен интелект (фиг. 9.2-2).

Машинното обучение (ML - Machine learning) е клас методи за решаване на проблеми на изкуствения интелект. Алгоритмите за машинно обучение разпознават модели в големи масиви от данни и ги използват, за да се самообучават. Всеки нов набор от данни позволява на математическите алгоритми да се усъвършенстват и адаптират в съответствие с получената информация, което позволява непрекъснато да се подобрява точността на препоръките и прогнозите.



**Фиг. 9.2-2 Избледняването на технологиите за създаване на данни и прилагането на инструменти за анализ отварят вратата към темата за машинното обучение.**

Както казва влиятелният главен изпълнителен директор на най-големия инвестиционен фонд в света (който притежава ключови дялове в почти всички най-големи компании за строителен софтуер, както и в компаниите, които притежават най-голямото количество недвижими имоти в света [55]) в интервю през 2023 г. - машинното обучение ще промени света на строителството.

*AI има огромен потенциал. Той ще промени начина, по който работим, и начина, по който живеем. ИИ и роботиката ще променят начина, по който работим, и начина, по който строим, и ще можем да използваме ИИ и роботиката като средство за създаване на много по-голяма производителност [153].*

- Изпълнителен директор на най-големия инвестиционен фонд в света, интервю, септември 2023 г.

Машинно обучение (ML) работи чрез обработка на големи количества данни, като използва статистически техники, за да имитира аспекти на човешкото мислене. Повечето компании обаче не разполагат с такива масиви от данни, а ако имат, те често не са достатъчно маркирани. Именно тук могат да помогнат семантичните технологии и трансферното обучение - техника, която позволява на ML да бъде по-ефективна при работа с малки количества данни, чиято осъществимост беше обсъдена в предишните глави на тази част.

Същността на трансферното учене се състои в това, че вместо да изучавате всяка задача от нулата, можете да използвате знания, придобити в свързани области. Необходимо е да се осъзнае, че моделите и откритията от други индустрии могат да бъдат адаптирани и приложени в строителния бранш. Например методите за оптимизиране на логистичните процеси, разработени в търговията на дребно, помагат за подобряване на ефективността на управлението на веригата за доставки в строителството. Анализът на големи масиви от данни, който се използва активно във финансите, може да се приложи за прогнозиране на разходите и управление на риска в строителните проекти. А технологиите за компютърно зрение и роботика, разработвани в промишлеността, вече намират приложение в автоматизирания контрол на качеството, наблюдението на безопасността и управлението на съоръженията на строителните обекти.

Трансферното обучение позволява не само да се ускори въвеждането на иновации, но и да се намалят разходите за тяхното разработване, като се използва вече натрупаният опит на други индустрии.

$$\text{labor productivity in construction} = f(\text{AI})$$

Фиг. 9.2-3 Изкуственият интелект технологиите и роботиката ще бъдат основната движеща сила на бъдещето за повишаване на производителността в строителната индустрия.

Човешкото мислене е организирано на подобен принцип: ние надграждаме придобитите преди това знания, за да решаваме нови проблеми (фиг. 4.4-19, фиг. 4.4-20, фиг. 4.4-21). В машинното обучение този подход също работи - като опростим модела на данните и го направим по-елегантен,

можем да намалим сложността на проблема за ML алгоритмите. Това от своя страна намалява необходимостта от големи количества данни и намалява изчислителните разходи.

## От субективна оценка към статистическа прогноза

Епохата, в която стратегическите решения зависеха от интуицията на отделните мениджъри (фигура 9.2-4), е в миналото. В условията на все по-конкурентна и предизвикателна икономическа среда субективният подход става твърде рискован и неефективен. Компаниите, които продължават да разчитат на личното си мнение, вместо обективно да анализират данни, губят способността си да реагират бързо на промените.

Конкурентната среда изисква точност и повторяемост въз основа на данни, статистически модели и изчислими вероятности. Решенията вече не могат да се основават на усещане, а трябва да се базират на корелации, тенденции и прогнозни модели, изведенни от анализи и машинно обучение. Това не е просто промяна в инструментите - това е промяна в логиката на мислене: от предположения към доказателства, от субективни вероятности към статистически изчислени отклонения, от усещане към факти.



**Фиг. 9.2-4 Ерата на решенията, вземани от HiPPO (мнението на най-високоплатения служител), ще остане в миналото с навлизането на големите данни и машинното обучение.**

Мениджърите, които са разчитали единствено на собствените си усещания, неизбежно ще се сблъскат с нова реалност: авторитетът вече не определя избора. В центъра на управлението сега

са системи, които анализират милиони параметри и вектори, идентифицират скрити модели и предлагат оптимални стратегии.

Основната причина, поради която компаниите днес все още избягват да прилагат ML, е липсата на прозрачност. Повечето модели работят като "черни кутии" за мениджърите, без да обясняват как точно стигат до заключенията си. Това води до проблеми: алгоритмите могат да затвърдят стереотипите и дори да създават хумористични ситуации, както в случая с чатбота на Microsoft, който бързо се превърна в токсичен инструмент за комуникация [154].

В "Дълбоко мислене" Гари Каспаров, бивш световен шампион по шахмат, разсъждава за поражението си от компютъра IBM Big Blue [155]. Той твърди, че истинската стойност на изкуствения интелект не е в копирането на човешкия интелект, а в допълването на нашите способности. ИИ трябва да изпълнява задачи, в които хората са слаби, а хората да внасят творчество. Компютрите промениха традиционния подход към анализа на шахмата. Вместо да създават увлекателни истории за партиите, компютърните шахматни програми оценяват всеки ход безпристрастно, само въз основа на действителната му сила или слабост. Каспаров отбелязва, че човешката склонност да разглежда събитията като последователни истории, а не като отделни действия, често води до погрешни заключения - не само в шаха, но и в живота като цяло.

Ето защо, ако планирате да използвате машинно обучение за прогнозиране и анализ, е важно да разберете основните му принципи - как работят алгоритмите и как се обработват данните - преди да започнете да използвате инструменти за машинно обучение и AI в работата си. Най-добрият начин да започнете работа е чрез практически опит.

Един от най-удобните инструменти за първоначално запознаване с темата за машинното обучение и прогнозирането е бележникът Jupyter и популярният класически набор от данни Titanic, който ще предостави визуално въведение в основните методи за анализ на данни и изграждане на ML модели.

## **Набор от данни "Титаник": "Здравей, свят" в света на анализите данни и големи данни**

Един от най-известните примери за използване на ML в анализа на данни е анализът на набора от данни за "Титаник", който често се използва за изследване на вероятността за оцеляване на пътниците. Изучаването на тази таблица е аналогично на програмата "Hello World" при изучаването на езици за програмиране.

При потъването на кораба "Титаник" през 1912 г. загиват 1502 от общо 2224 души. Наборът от данни за "Титаник" съдържа не само информация за това дали даден пътник е оцелял, но и атрибути като: възраст, пол, клас на билета и други параметри. Този набор от данни е достъпен безплатно и може да бъде отворен и анализиран в различни онлайн платформи.

Връзка към набора от данни за Титаник:

<https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv>

По-рано в главата "IDE с поддръжка на LLM и бъдещи промени в програмирането" вече обсъдихме

Jupyter Notebook - една от най-популярните среди за разработка за анализ на данни и машинно обучение. Безплатни облачни аналоги на Jupyter Notebook са платформите Kaggle и Google Collab, които ви позволяват да изпълнявате код на Python без инсталација на софтуер и осигуряват безплатен достъп до изчислителни ресурси.

Kaggle е най-голямата платформа за състезания по анализ на данни и машинно обучение с интегрирана среда за изпълнение на код. Към октомври 2023 г. Kaggle има над 15 милиона потребители [156] от 194 държави.

Изтеглете и използвайте набора от данни на Titanic в платформата Kaggle (Фигура 9.2-5), за да съхранявате набора от данни (негово копие) и да изпълнявате кода на Python с предварително инсталирани библиотеки директно в браузъра, без да се налага да инсталirate специален IDE.



Фиг. 9.2-5 Статистически данни за таблицата на Титаник - най-популярният набор от данни за обучение за анализ на данни и машинно обучение.

Наборът от данни за "Титаник" включва данни за 2224 пътници на борда на кораба RMS *Titanic* по време на корабокрушението му през 1912 г. Наборът от данни е представен като две отделни таблици - обучителна (train.csv) и тестова (test.csv) извадка, което позволява да се използва както за обучение на модели, така и за оценка на тяхната точност върху нови данни.

Наборът от данни за обучение съдържа както атрибути - атрибути на пътниците (възраст, пол, клас билети и други), така и информация за това кой е оцелял (колона с двоични стойности "Оцелял"). Наборът от данни за обучение (фигура 9.2-6 - файл train.csv) се използва за обучение на модела. Тестовият набор от данни (фигура 9.2-7 - файл test.csv) включва само атрибути на пътниците без информация за оцелелите (без нито една колона "Оцелял"). Тестовият набор от данни е предназначен за тестване на модела върху нови данни и за оценка на неговата точност.

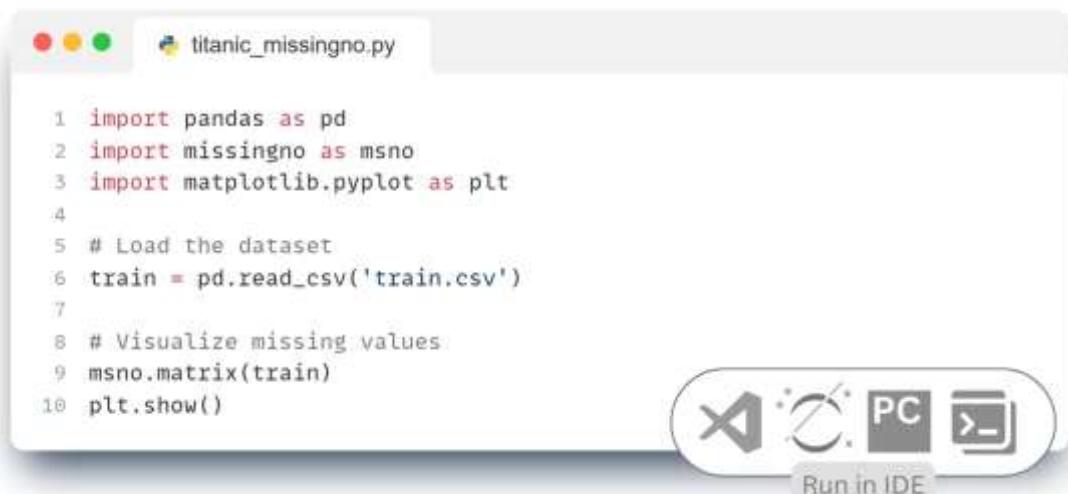
По този начин имаме почти идентични атрибути на пътниците в набора от данни за обучение и набора от данни за тестване. Единствената основна разлика е, че в тестовото множество от данни

имаме списък с пътници, които нямат колона "Оцелял" - целевата променлива, която искаме да се научим да предсказваме с помощта на различни математически алгоритми. А след като изградим модела, ще можем да сравним изхода на нашия модел с реалния параметър "Survivor" от тестовия набор от данни, който ще вземем предвид, за да оценим резултатите.

Основни колони на таблицата, параметри на пътниците в набора от данни за обучение и набор от данни за изпитване:

- **PassengerId** - уникален идентификатор на пътника
- **Оцелял** - 1, ако пътникът е оцелял, 0, ако е загинал (не е налично в тестовия набор)
- **Pclass** - клас на билета (1, 2 или 3)
- **Име** - името на пътника
- **Пол** - пол на пътника (мъж/жена)
- **Възраст**
- **SibSp** - брой братя/сестри или съпрузи на борда
- **Parch** - брой родители или деца на борда
- **Билет** - номер на билета
- **Такса** - цена на билета
- **Кабина** - номер на кабината (липсват много данни)
- **Embarked** е пристанището на качване (C = Cherbourg, Q = Queenstown, S = Southampton).

За да визуализирате липсващите данни в двете таблици, можете да използвате библиотеката missingno (фиг. 9.2-6, фиг. 9.2-7), която показва липсващите стойности под формата на хистограма, където белите полета показват липсващи данни. Тази визуализация позволява бърза оценка на качеството на данните преди обработката им.

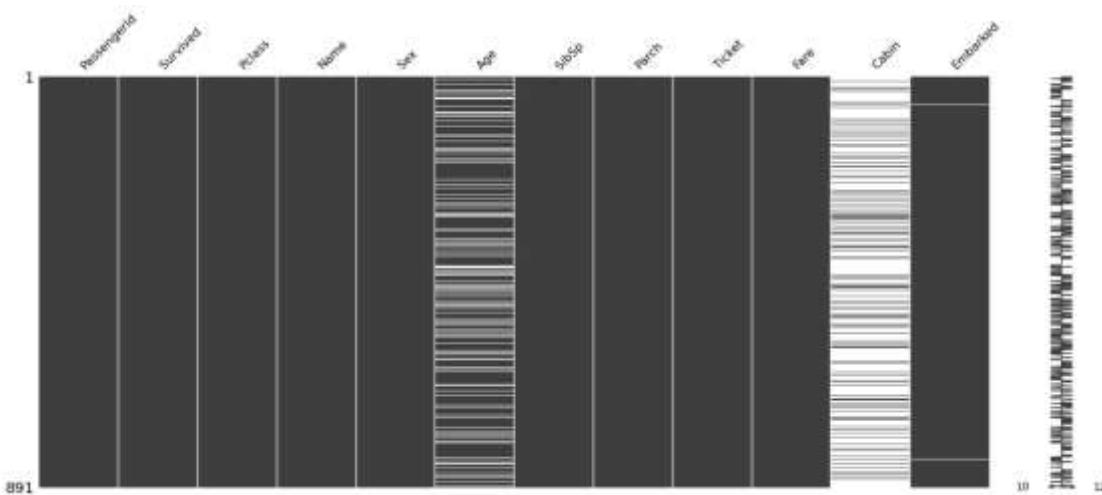


```

1 import pandas as pd
2 import missingno as msno
3 import matplotlib.pyplot as plt
4
5 # Load the dataset
6 train = pd.read_csv('train.csv')
7
8 # Visualize missing values
9 msno.matrix(train)
10 plt.show()

```

The screenshot shows a Jupyter Notebook cell with the title "titanic\_missingno.py". The cell contains Python code for visualizing missing values in a dataset. The code imports pandas, missingno, and matplotlib.pyplot, loads a "train.csv" dataset, and then uses the missingno library's matrix function to visualize missing values. A "Run in IDE" button is visible at the bottom right of the cell.



Фиг. 9.2-6 Няколко реда код се използват за визуализиране на липсващи данни в набора от данни за обучение на Titanic, където ключовият параметър за обучение е параметърът "Оцелял".



Фигура 9.2-7 Визуализация на липсващи данни в тестовия набор от данни за Titanic, който съдържа само характеристики на пътниците без информация.

Преди да се формулират хипотези и да се направят прогнози въз основа на набора от данни, визуалните анализи помагат да се идентифицират ключови модели в данните, да се оцени тяхното качество и да се определят възможни зависимости. Съществуват много техники за визуализация, които могат да ви помогнат да разберете по-добре набора от данни за "Титаник". Можете да използвате диаграми на разпределението, за да анализирате възрастовите групи на пътниците, диаграми на оцеляването по пол и класа, както и матрици на липсващите данни, за да оцените качеството на информацията и да разберете данните.

- ❷ Нека да помолим LLM да ни помогне да визуализираме данните от набора от данни за Титаник, като изпратим следната текстова заявка до всеки модел на LLM (CHATGP, LLaMa, Mistral DeepSeek, Grok, Claude, QWEN или друг):

Моля, покажете няколко прости графики за набора от данни за Титаник. Изтеглете сами набора от данни и покажете ↵

- ❸ LLM отговор под формата на готов код и графики, визуализиращи параметрите на набора от данни

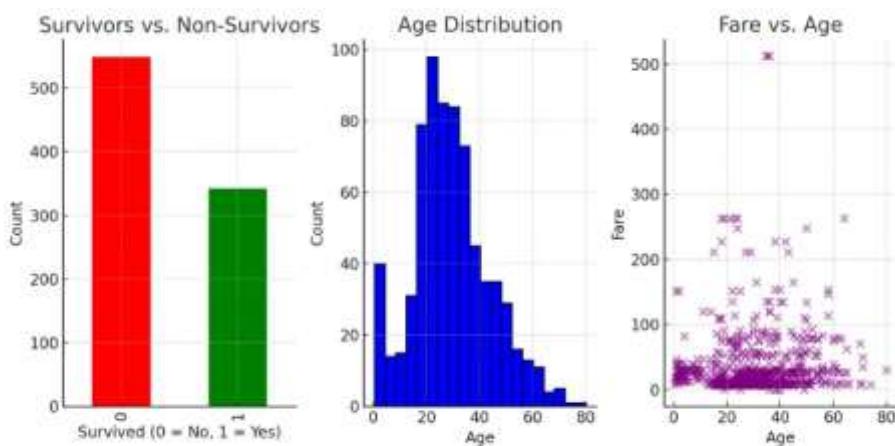
Create in LLM Chat

```

1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 # Load the dataset
5 file_path = "/mnt/data/titanic.csv"
6 df = pd.read_csv(file_path)
7
8 # Convert column names to lowercase for consistency
9 df.columns = df.columns.str.lower()
10
11 # Basic Plots
12 plt.figure(figsize=(10, 5))
13
14 # Bar Chart: Survival Count
15 plt.subplot(1, 3, 1)
16 df['survived'].value_counts().plot(kind='bar', color=['red', 'green'])
17 plt.title("Survivors vs. Non-Survivors")
18 plt.xlabel("Survived (0 = No, 1 = Yes)")
19 plt.ylabel("Count")
20
21 # Histogram: Age Distribution
22 plt.subplot(1, 3, 2)
23 df['age'].dropna().plot(kind='hist', bins=20, color='blue', edgecolor='black')
24 plt.title("Age Distribution")
25 plt.xlabel("Age")
26 plt.ylabel("Count")
27
28 # Scatter Plot: Fare vs. Age
29 plt.subplot(1, 3, 3)
30 plt.scatter(df['age'], df['fare'], alpha=0.5, color='purple')
31 plt.title("Fare vs. Age")
32 plt.xlabel("Age")
33 plt.ylabel("Fare")
34
35 # Show the plots
36 plt.tight_layout()
37 plt.show()

```

Run in IDE



Фигура 9.2-8 LLM ви помага да получите незабавна визуализация на данните от вашия набор от данни.

Визуализацията на данните е важна стъпка за подготовка на набора от данни за последващото изграждане на модел за машинно обучение, до който може да се стигне само чрез разбиране на данните.

## Машинно обучение в действие: от пътниците на "Титаник" до управлението на проекти

Основната хипотеза, използвана за изследване на рамката за машинно обучение въз основа на набора от данни за "Титаник", е, че определени групи пътници са имали по-голям шанс да оцелеят.

Малката таблица на пътниците от "Титаник" стана популярна в целия свят и милиони хора я използват за обучение, експериментиране и тестване на модели, за да разберат кои алгоритми и хипотези ще създадат най-точния модел за прогнозиране на оцеляването въз основа на набора от данни за обучение на пътниците от "Титаник".

Привлекателността на набора от данни за "Титаник" се дължи на неговата компактност: с няколкостотин реда и дванадесет колони (фиг. 9.2-6) той предоставя широки възможности за анализ. Наборът от данни е, сравнително просто, класически пример за решение за бинарна класификация, където целта на проблема - оцеляване - е изразена в удобния формат 0 или 1.

Джон Уилър в "It from Bit" [7] твърди, че Вселената се основава на двоичен избор. По подобен начин бизнесът, управляван от хора, съставени от молекули, всъщност е изграден върху поредица от бинарни двоични избори.

Освен това данните се основават на реално историческо събитие, което ги прави ценни за изследвания, за разлика от изкуствено създадените примери. Само на платформата Kaggle, една от най-големите Data Pipeline и ETL, 1 355 998 души са участвали в предизвикателствата, базирани на набора от данни за Титаник, като са разработили 53 963 уникални решения за Data Pipeline [157] (фигура 9.2-9).

Изглежда невероятно, но само 1000 реда данни за пътниците на "Титаник" с 12 параметъра се превърнаха в поле за милиони хипотези, логически вериги и уникални Data Pipelines. От един малък набор от данни се раждат безкрайни прозрения, хипотези и интерпретации - от прости модели за оцеляване до сложни ансамбли, които отчитат скрити закономерности и сложни лабиринти от разсъждения.

**Machine Learning from Disaster**

**Submit Prediction**

Data    Code    Models    Discussion    Leaderboard    Rules

**Titanic Tutorial**    Updated 3y ago    Score: 0.80143 · 29858 comments · Titanic - Machine Learning from Disaster    16916 Gold

**Titanic competition w/ TensorFlow Decision Forests**    Updated 2y ago    Score: 0.80143 · 318 comments · Titanic - Machine Learning from Disaster    1098 Gold

**Titanic Data Science Solutions**    Updated 6y ago    Score: 0.80382 · 2590 comments · Titanic - Machine Learning from Disaster    10723 Gold

**Exploring Survival on the Titanic**    Updated 7y ago    Score: 0.80382 · 1072 comments · Titanic - Machine Learning from Disaster    3968 Gold

**Фиг. 9.2-9** Първите пет решения от общо 53 963 готови решения и решения с отворен код на Pipeline. Почти 1,5 miliona души вече са се опитали да решат този проблем само в Kaggle [157].

Ако дори такава малка таблица може да генерира милиони уникални решения (фиг. 9.2-9), какво можем да кажем за реални набори от данни за промишлено строителство, където параметрите се измерват в десетки хиляди?

Един стандартен CAD -проект на сравнително малка сграда съдържа десетки хиляди обекти с хиляди параметри - от геометрични характеристики до разходи и времеви характеристики. Представете си колко много потенциални прозрения, връзки, прогнози и хипотези за управление се крият в данните от всички проекти на вашата компания, събиращи през годините. Историческите данни за проектите не са просто архив - те са живата памет на организацията, нейният цифров отпечатък, който може да бъде анализиран, за да се изградят голям брой уникални хипотези.

Най-важното е, че не е необходимо да чакате общността на Kaggle да прояви интерес към вашата компания или вашите данни. Можете да започнете да работите с това, с което разполагате още днес: да извършвате анализи на собствените си данни, да обучавате модели на собствените си данни, да идентифицирате повторения, аномалии и модели. Там, където преди се изискваха години на експериментиране и скъпи консултации, сега са достатъчни инициативност, магистратура, отворен подход към данните и желание да се учате.

- ❷ За да създадем алгоритъм за машинно обучение, който ще предсказва степента на оцеляване на пътниците въз основа на набора от данни за обучение на пътници train.csv, нека помолим LLM да реши този проблем вместо нас:

Въз основа на набор от данни за обучение на пътниците от "Титаник" създайте модел за машинно обучение за прогнозиране на оцеляването. ↵

2 Отговор на LLM :

The diagram illustrates a vertical workflow for building a machine learning model. At the top is a computer monitor icon with code labeled 'titanic\_predict.py'. A downward arrow leads to a document icon labeled 'train.csv'. Another downward arrow leads to a laptop icon with a gear and brain, representing the machine learning process.

```

1 import pandas as pd
2 import seaborn as sns
3 import matplotlib.pyplot as plt
4 from sklearn.model_selection import train_test_split
5 from sklearn.ensemble import RandomForestClassifier
6 from sklearn.metrics import accuracy_score

1 # Load the dataset
2 train = pd.read_csv('train.csv')

1 # Convert categorical data
2 train['Sex'] = train['Sex'].map({'male': 0, 'female': 1})
3 train = train.dropna(subset=['Age']) # Remove missing age values
4
5 # Select relevant features
6 features = ["Pclass", "Sex", "Age", "SibSp", "Parch", "Fare"]
7 X = train[features]
8 y = train['Survived']
9
10 # Split the dataset into training and testing sets
11 X_train, X_test, y_train, y_test = train_test_split(X, y,
12 test_size=0.2, random_state=42)
13
14 # Train the model
15 model = RandomForestClassifier(n_estimators=100, random_state=42)
16 model.fit(X_train, y_train)
17
18 # Make predictions
19 y_pred = model.predict(X_test)
20
21 # Evaluate accuracy
22 accuracy = accuracy_score(y_test, y_pred)
23 print(f'Model Accuracy: {accuracy:.2f}')

```

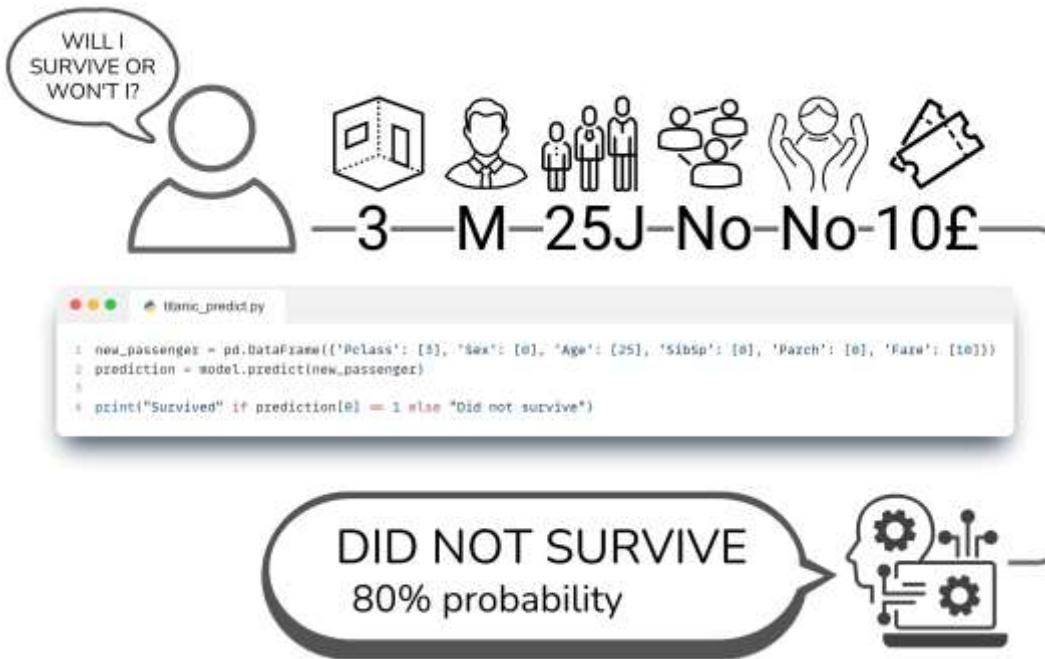
Фиг. 9.2-10 LLM построи прогноза за оцелелите от Титаник, като използва алгоритъма за машинно обучение Random Forest.

Полученият код от LLM (фиг. 9.2-10) зарежда данни за пътниците на Титаник, почиства ги, преобразува категорични променливи (например пол в цифров формат) и обучава модела чрез алгоритъма RandomForestClassifier, за да предскаже дали даден пътник е оцелял или не (ще говорим повече за популярните алгоритми в следващите глави).

Кодът разделя данните за обучение на набори за обучение и тестови набори (на уебсайта на Kaggle вече са създадени готови набори test.csv (фиг. 9.2-7) и train.csv (фиг. 9.2-6) за обучение, след което моделът се обучава върху данните за обучение и се тества върху тестовите данни, за да се види колко добър е даден модел за предсказване. След обучението тестовите данни от test.csv (с реални данни за тези, които са оцелели или не са оцелели) се подават на модела и той предсказва кой е оцелял и кой не. В нашия случай точността на модела за машинно обучение, който получихме, е около 80%, което показва, че той улавя моделите доста добре.

Машинното обучение може да се сравни с дете, което се опитва да постави правоъгълно блокче в кръгла дупка. В началните етапи алгоритъмът опитва много подходи, като се натъква на грешки и несъответствия. Този процес може да изглежда неефективен, но осигурява важно обучение: анализирали всяка грешка, моделът подобрява прогнозите си и взема все по-точни решения.

Сега този модел (фиг. 9.2-10) може да се използва за прогнозиране на преживяемостта на новите пътници и например, ако го захраните с информация за пътниците, като използвате функцията model.predict с параметрите: "мъж", "3-та класа", "25 години", "няма роднини на борда", моделът ще даде прогноза - че пътник с 80% вероятност няма да оцелее след катастрофата, ако е бил на "Титаник" през 1912 г. (фиг. 9.2-11).



**Фигура 9.2-11** Моделът, който създадохме по-горе, сега може да предскаже с 80% вероятност дали някой нов пътник на "Титаник" ще оцелее.

Моделът за прогнозиране на оцеляването на пасажерите на "Титаник" илюстрира много по-широка концепция: хиляди професионалисти в строителния бранш всеки ден вземат подобни "двойни" решения - животът или смъртта на дадено решение, проект, оценка, инструмент, печалба или загуба, безопасност или рисък. Както в примера с "Титаник", където изходът зависеше от фактори (пол,

възраст, класа), в строителството всеки аспект на решението се влияе от много собствени фактори и променливи (колони от таблици): цена на материалите, квалификация на работниците, крайни срокове, време, логистика, технически рискове, коментари и стотици хиляди други параметри.

В строителния бранш машинното обучение следва същите принципи като в други области: моделите се обучават на исторически данни - от проекти, договори, оценки - за да се тестват различни хипотези и да се намерят най-ефективните решения. Този процес много прилича на обучението на дете чрез проби и грешки: с всеки цикъл моделите се адаптират и стават по-точни.

Използването на натрупани данни открива нови хоризонти пред строителството. Вместо отнемащите време ръчни изчисления могат да се обучат модели, които могат да прогнозират ключови характеристики на бъдещи проекти с висока степен на точност. По този начин прогнозният анализ превръща строителната индустрия в пространство, в което може не само да се планира, но и уверено да се прогнозира развитието.

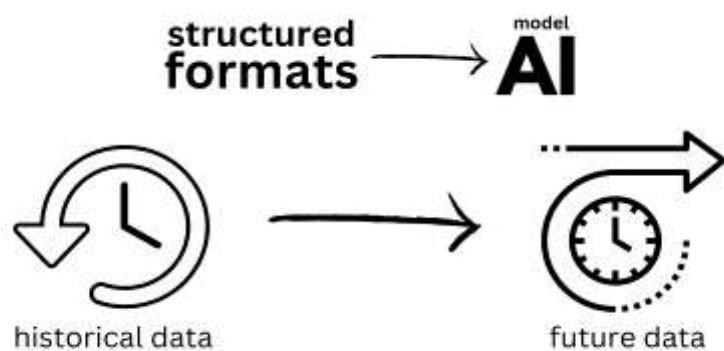
## **Предвиждания и прогнози въз основа на исторически данни**

Данните, събрани за проектите на компанията, дават възможност за създаване на модели, които могат да прогнозират разходните и времевите характеристики на бъдещи, все още нереализирани обекти - без времеемки ръчни изчисления и сравнения. Това дава възможност значително да се ускорят и опростят процесите на оценяване, като се разчита не на субективни предположения, а на обосновани математически прогнози.

По-рано, в четвъртата част на книгата, разглеждахме подробно традиционните методи за оценка на разходите по проекти, включително метода, основан на ресурсите, а също така споменахме параметричните и експертните подходи. Тези методи все още са актуални, но в съвременната практика те започват да се обогатяват с инструменти за статистически анализ и машинно обучение, които могат значително да подобрят точността и възпроизвеждането на оценките.

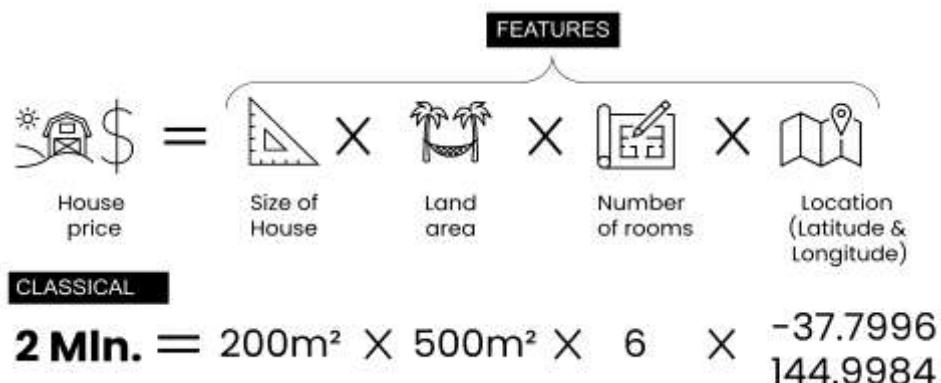
Процесите на ръчно и полуавтоматично изчисляване на цени и времеви атрибути в бъдеще ще бъдат допълнени от мнението и прогнозите на ML модели, способни да анализират исторически данни, да откриват скрити модели и да предлагат информирани решения. Нови данни и сценарии ще се генерираат автоматично от вече наличната информация - подобно на начина, по който езиковите модели (LLM) създават текстове, изображения и кодове въз основа на данни, събиращи през годините от отворени източници [158].

Както днес хората разчитат на опита, интуицията и вътрешната статистика, за да оценят бъдещи събития, така през следващите години бъдещето на строителните проекти все повече ще се определя от комбинация от натрупани знания и математически модели за машинно обучение.



**Фиг. 9.2-12** Качествените и структурирани исторически данни за компанията са материалът, върху който се изграждат моделите за машинно обучение и прогнозите.

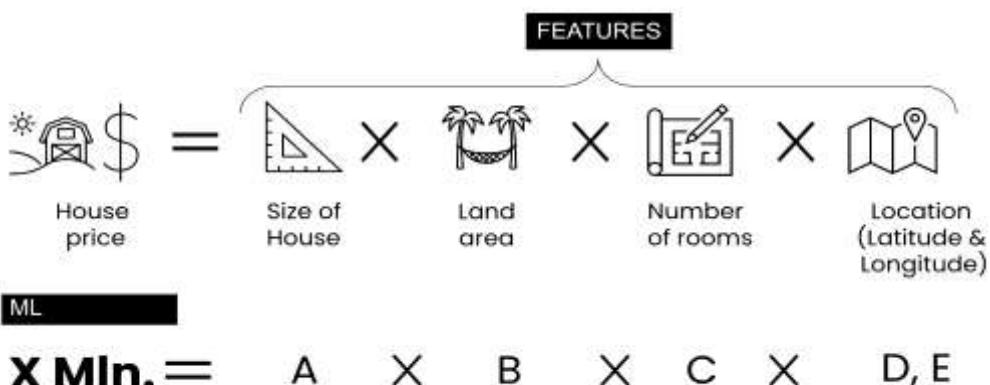
Да разгледаме един прост пример: прогнозиране на цената на къща въз основа на нейната площ, размер на парцела, брой стаи и географско местоположение. Един от подходите е да се изгради класически модел, който анализира тези параметри и изчислява очакваната цена (фигура 9.2-13). Този подход изиска точна и предварително известна формула, което е практически невъзможно в реалната практика.



**Фиг. 9.2-13** Класически алгоритъм може да се използва за оценка на стойността на къща с фиксирана формула, която трябва да се намери.

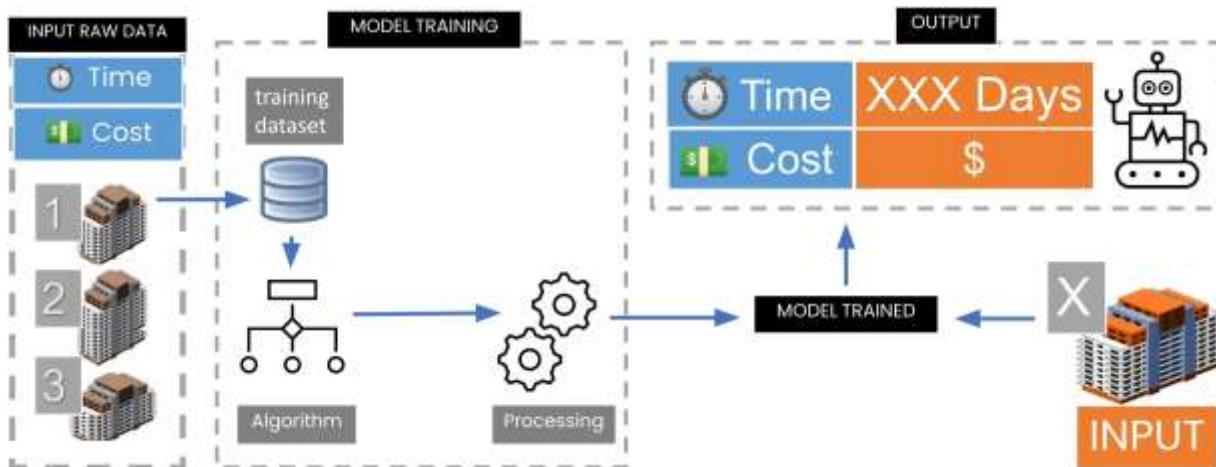
Машинното обучение елиминира ръчното търсене на формули и ги заменя с обучени алгоритми, които самостоятелно идентифицират зависимости, многократно по-точни от всички предварително зададени уравнения. Алтернативно, нека създадем алгоритъм за машинно обучение, който ще генерира модел въз основа на предварително разбиране на проблема и исторически данни, които може да са непълни (фигура 9.2-14).

Ако използваме ценообразуването като пример, машинното обучение позволява създаването на различни видове математически модели, които не изискват познаване на точния механизъм на формиране на разходите. Моделът се "учи" от данните за предишни проекти, като се приспособява към реалните закономерности между параметрите на сградите, тяхната цена и крайните срокове.



Фиг. 9.2-14 За разлика от класическото оценяване по формула, алгоритъмът за машинно обучение се обучава на исторически данни.

В контекста на контролираното машинно обучение, всеки проект в набора от данни за обучение съдържа както входни атрибути (напр. разходи и време на подобни сгради), така и очаквани изходни стойности (напр. разходи или време). Подобен набор от данни се използва за създаване и адаптиране на модел за машинно обучение (фигура 9.2-15). Колкото по-голям е наборът от данни и колкото по-високо е качеството на данните в него, толкова по-точен ще бъде моделът и толкова по-точни ще бъдат резултатите от прогнозите.



Фигура 9.2-15 Един ML модел, обучен на данни за разходите и графика от минали проекти, ще определи разходите и графика на нов проект с определена вероятност.

След като моделът е създаден и обучен да оценява строителството на нов проект, просто предоставете на модела нови атрибути за новия проект и моделът ще предостави приблизителни резултати въз основа на предварително научени модели с определена вероятност.

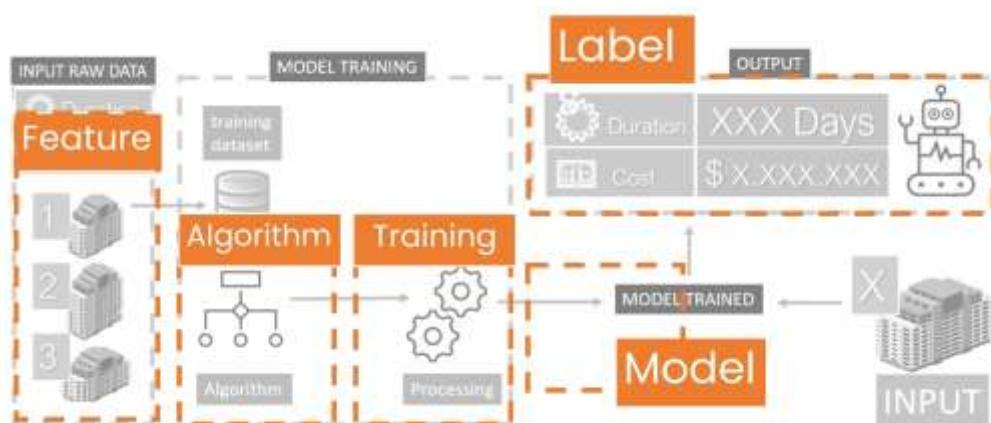
## Ключови концепции за машинно обучение

Машинното обучение не е магия, а просто математика, данни и откриване на модели. То няма ре-

лен интелект, а е програма, обучена на данни, която разпознава модели и взема решения без постоянно участие на човека.

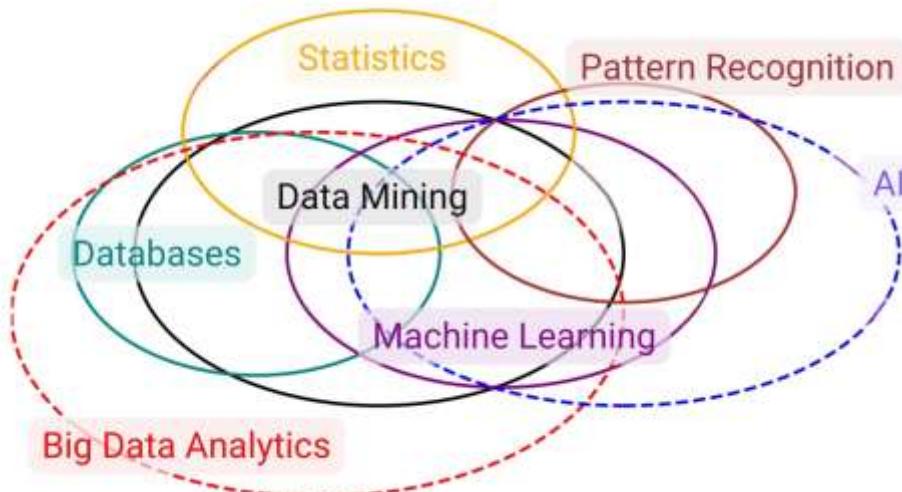
Машинното обучение използва няколко ключови понятия за описание на своята структура (Фигура 9.2-16):

- **Етикетите** са целеви променливи или атрибути (параметърът "Survivor" (оцелял) в набора от данни за "Титаник"), които моделът трябва да предвиди. Пример: разходи за строителство (напр. в долари), продължителност на строителните работи (напр. в месеци).
- **Характеристиките** са независими променливи или атрибути, които служат като входни данни за модела. В модела за прогнозиране те се използват за прогнозиране на етикети. Примери: размер на парцела (в квадратни метри), брой етажи на сградата, обща площ на сградата (в квадратни метри), географско местоположение (географска ширина и дължина), вид на материалите, използвани в строителството. Броят на характеристиките определя и размерността на данните.
- **Моделът** представлява набор от различни хипотези, една от които апроксимира целевата функция, която трябва да бъде предсказана или апроксимирана. Пример: модел за машинно обучение, който използва техники за регресионен анализ за прогнозиране на разходите и сроковете за строителство.
- **Алгоритъм за обучение** Алгоритъмът за обучение е процесът на намиране на най-добрата хипотеза в модела, която точно съответства на целевата функция, като се използва набор от данни за обучение. Пример: Алгоритъм за линейна регресия, KNN или случайна гора, който анализира данни за разходите и времето за строителство, за да идентифицира връзки и модели.
- **Обучение** - По време на процеса на обучение алгоритъмът анализира данните за обучение, като намира модели, които съответстват на връзката между входните атрибути и целевите етикети. Резултатът от този процес е обучен модел за машинно обучение, готов за прогнозиране. Пример: процес, при който алгоритъмът анализира исторически данни за строителството (разходи, време, характеристики на обекта), за да създаде прогнозен модел.



Фигура 9.2-16 ML използва етикети и атрибути за създаване на модели, които се обучават върху данни с помощта на алгоритми за прогнозиране на резултати.

Машинното обучение не съществува изолирано, а е част от по-широва екосистема от аналитични дисциплини, включително статистика, бази данни, извличане на данни, разпознаване на образи, анализ на големи обеми от данни и изкуствен интелект. Фигура 9.2-17 показва как тези области се припокриват и допълват взаимно, осигурявайки цялостна рамка за съвременните системи за вземане на решения и автоматизация.



Фиг. 9.2-17 Връзката между различните области на анализа на данни: статистика, машинно обучение, изкуствен интелект, големи масиви от данни, разпознаване на образи и извличане на данни.

Основната цел на машинното обучение е да предостави на компютрите способността автоматично да усвояват знания без човешка намеса или помощ и да коригират съответно своите действия [159].

Така че в бъдеще ролята на човека ще бъде само да предоставя на машината познавателни способности - той ще задава условията, тежестите и параметрите, а моделът за машинно обучение ще свърши останалото.

В следващата глава ще разгледаме конкретни примери за прилагане на алгоритми. Ще бъдат използвани реални таблици и опростени модели, за да се покаже как се изгражда прогнозата стъпка по стъпка.



## ГЛАВА 9.3.

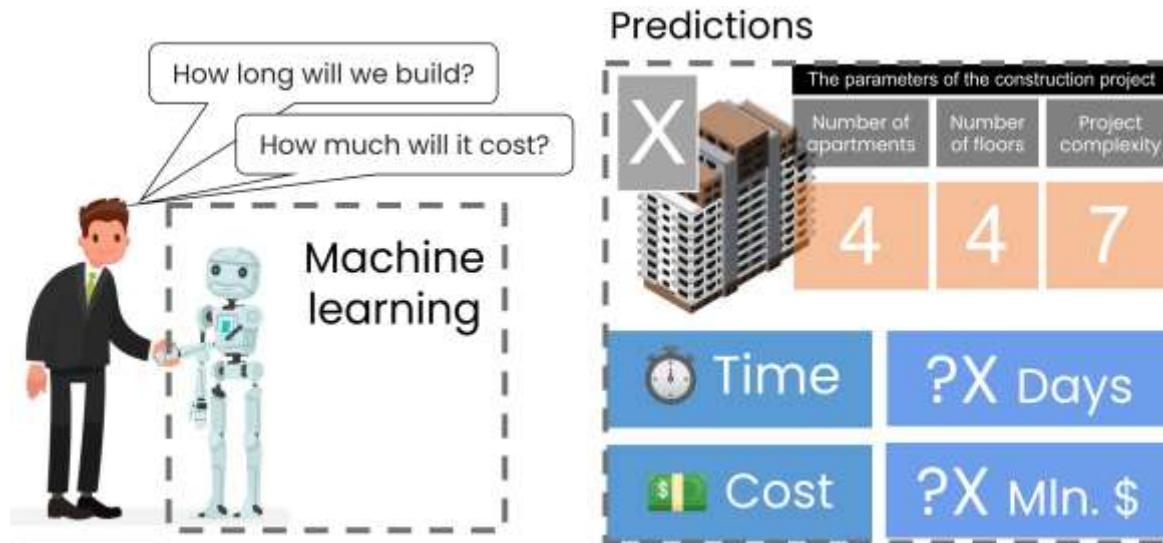
### ПРОГНОЗИРАНЕ НА РАЗХОДИ И СРОКОВЕ С ПОМОЩТА НА МАШИННО ОБУЧЕНИЕ

#### Пример за използване на машинно обучение за намиране на разходите и сроковете на проекта

Оценката на времето и разходите за строителство е един от ключовите процеси в дейността на строителната компания. Традиционно тези оценки се правят от експерти въз основа на опит, спрачници и нормативни бази данни. С цифровата трансформация и нарастващата наличност на данни обаче вече е възможно да се използват модели за машинно обучение (ML) за подобряване на точността и автоматизиране на такива оценки.

Въвеждането на машинно обучение в процеса на изчисляване на разходите и сроковете за строителство не само позволява по-ефективно планиране, но и се превръща в отправна точка за интегриране на интелигентни модели в други бизнес процеси - от управление на риска до оптимизиране на логистиката и снабдяването.

Важно е да можете бързо да определите колко време ще отнеме изграждането на даден проект и каква ще бъде общата му стойност. Тези въпроси за времето и цената на проекта традиционно са били на преден план в съзнанието както на клиентите, така и на строителните компании още от зараждането на строителната индустрия.



Фиг. 9.3-1 В строителните проекти бързината и качеството на оценяване на времето и разходите за строителство са ключови фактори за успех.

В следващия пример ще бъдат извлечени ключови данни от минали проекти и ще бъдат използвани за разработване на модел за машинно обучение, което ще ни позволи да използваме модела за оценка на разходите и сроковете на нови строителни проекти с нови параметри (фигура

9.3-1).

Разгледайте три проекта с три ключови характеристики: брой апартаменти (където 100 апартамента са равностойни на числото 10 за по-лесна визуализация), брой етажи и условна мярка за сложност на строителството по скала от 1 до 10, където 10 е най-високата оценка за сложност. В машинното обучение процесът на преобразуване и опростяване на стойности като 100 към 10 или 50 към 5 се нарича "нормализация".

**Нормализирането** в машинното обучение е процесът на привеждане на различни цифрови данни към обща скала, за да се улесни обработката и анализът. Този процес е особено важен, когато данните имат различни мащаби и мерни единици.

Да предположим, че първият проект (фигура 9.3-2) има 50 апартамента (след нормализация - 5), 7 етажа и оценка за сложност 2, което означава сравнително просто строителство. Вторият проект вече е имал 80 апартамента, 9 етажа и относително сложна конструкция. При тези условия строителството на първата и втората жилищна сграда е отнело 270 и 330 дни, а общата стойност на проекта е била съответно 4,5 млн. долара и 5,8 млн. долара.

Construction project	The parameters of the construction project			The key parameters of the project	
	Number of apartment	Number of floors	Project complexity	Time	Cost
1	50 = 100/2	7	2	270	\$ 4.502.000
2	80 = 100*8/10	9	6	330	\$ 5.750.000
3	30 = 100*3/10	5	3	230	\$ 3.262.000
X	40 = 100*4/10	4	7	?X	\$ ?X. XXX.XXX

Фигура 9.3-2 Пример за набор от минали проекти, които ще се използват за оценка на времето и разходите на бъдещия проект X.

Когато се изгражда модел за машинно обучение за такива данни, основната задача е да се идентифицират критичните атрибути (или етикети) за прогнозиране, в този случай времето за строителство и разходите. При малък набор от данни ще използваме информация за предишни строителни проекти, за да планираме нови: използвайки алгоритми за машинно обучение, трябва да предвидим разходите и продължителността на строителството на нов проект X въз основа на дадени атрибути на новия проект, като например 40 апартамента, 4 етажа и относително висока сложност на проекта 7 (фиг. 9.3-2). В реални условия броят на входните параметри може да бъде много по-голям, като варира от няколко десетки до стотици фактори. Те могат да включват: вид на строителните материали, климатична зона, ниво на квалификация на изпълнителите, наличие на комунални услуги, вид на основите, сезон на започване на работите, коментари на майсторите и др.

За да създадем прогнозен модел за машинно обучение, трябва да изберем алгоритъм за създаването му. Алгоритъмът в машинното обучение е като математическа рецепта, която учи компютъра как да прави прогнози (да смесва в правилния ред параметрите) или да взема решения въз основа на данни.

За да се анализират данни за минали строителни проекти и да се прогнозират времето и разходите за бъдещи проекти (фигура 9.3-2), може да се използва един популярен алгоритъм за машинно обучение:

- **Линейна регресия (Линейна регресия):** този алгоритъм се опитва да намери пряка връзка между атриутите, например между броя на етажите и разходите за строителство. Целта на алгоритъма е да се намери линейно уравнение, което най-добре описва тази връзка, което позволява да се правят прогнози.
- **Алгоритъм k-най-близки съседи (k-NN):** този алгоритъм сравнява нов проект с предишни проекти, които са били сходни по размер или сложност. k-NN класифицира данните въз основа на това кой от k (броя) обучителни примери е най-близък до тях. В контекста на регресията резултатът е средната стойност или медианата на k най-близки съседи.
- **Дървета за вземане на решения:** това е модел за прогнозно моделиране, който разделя данните на подмножества въз основа на различни условия, като използва дървовидна структура. Всеки възел на дървото представлява условие или въпрос, водещ до по-нататъшно разделяне на данните, а всяко листо представлява крайната прогноза или резултат. Алгоритъмът разделя данните на по-малки групи въз основа на различни характеристики, например първо по брой истории, след това по сложност и т.н., за да направи прогноза.

Нека разгледаме алгоритмите за машинно обучение за оценка на разходите за нов проект, като използваме за пример два популярни алгоритъма: линейна регресия и алгоритъма на най-близките съседи.

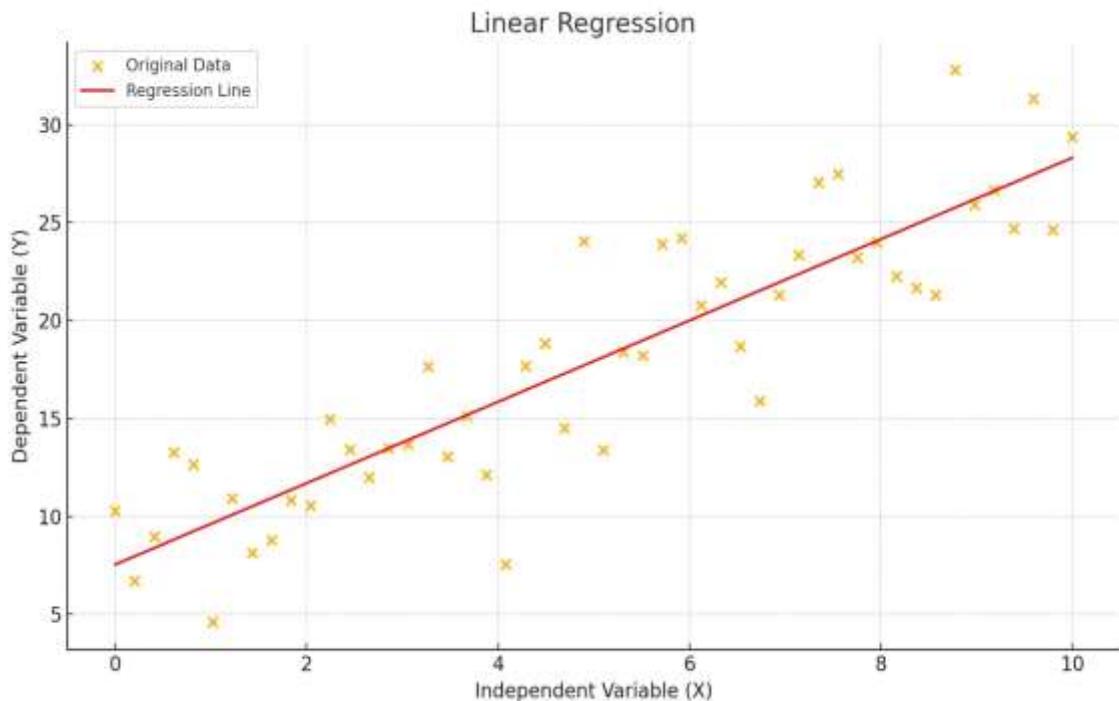
## Прогнозиране на разходите и времето за изпълнение на проекта чрез линейна регресия

Линейна регресия е основен алгоритъм за анализ на данни, който предвижда стойността на дадена променлива въз основа на линейна връзка с една или повече други променливи. Този модел предполага, че съществува пряка линейна връзка между зависимата променлива и една или повече независими променливи, а целта на алгоритъма е да намери тази връзка.

Простотата и яснотата на линейната регресия я направиха популяррен инструмент в различни области. Когато се работи с една променлива, линейната регресия се състои в намирането на най-подходящата линия през точките с данни.

Линейната регресия намира най-добрата права линия (червена линия), която апроксимира зависимостта между входната променлива X и изходната променлива Y. Тази линия позволява да се про-

гнозират стойностите на Y за нови стойности на X въз основа на идентифицираната линейна зависимост (фигура 9.3-3).



**Фиг. 9.3-3 Принципът на линейната регресия е да се намери най-добрата права линия през обучаващите стойности.**

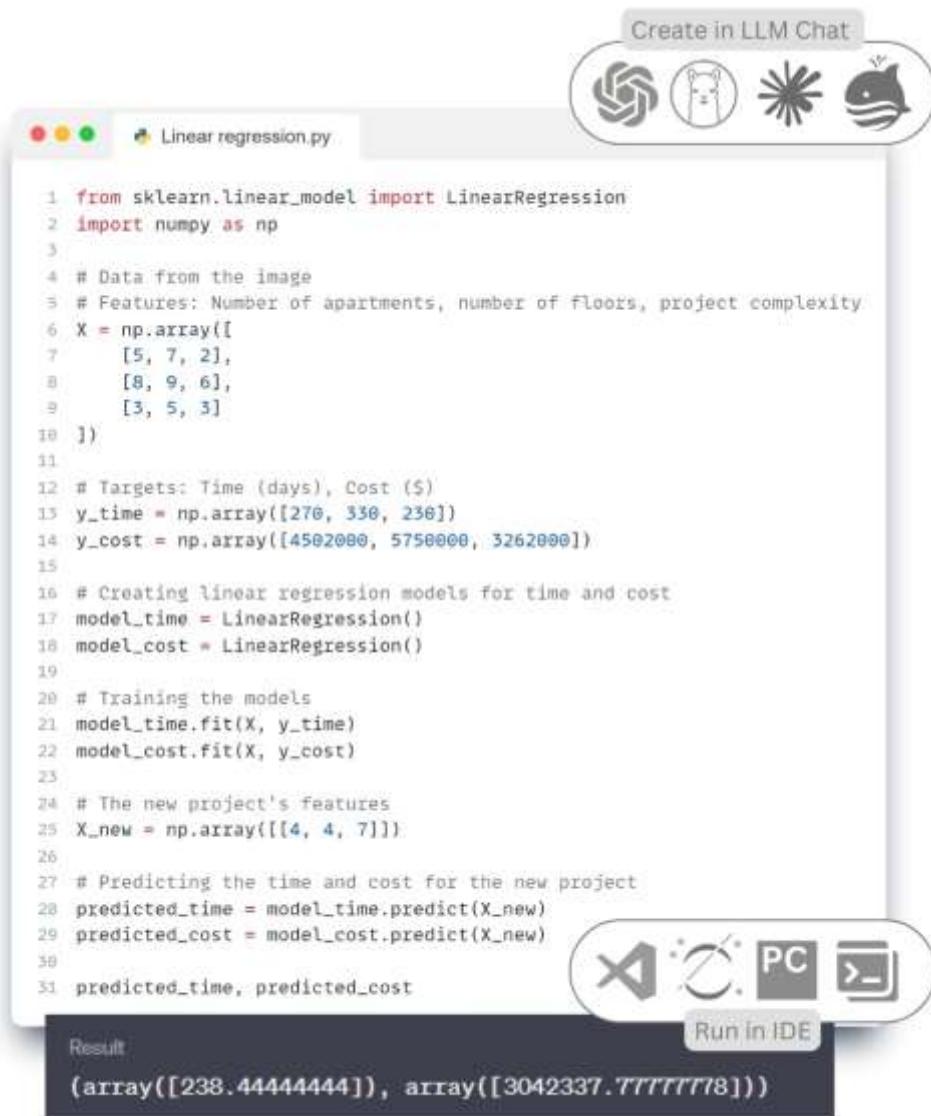
Тази линия е представена чрез уравнение, в което въвеждането на стойност на независимата променлива (X) води до прогнозиране на стойността на зависимата променлива (Y). Този процес позволява ефективно прогнозиране на Y на базата на известни стойности на X, като се използва линейна зависимост между тях. Пример за намиране на такава статистически осреднена линия може да се види в оценката на данните за разрешителните за строеж в Сан Франциско (фигура 9.1-7), където инфлацията с помощта на линейна регресия е изчислена за различни видове съоръжения.

Нека да заредим таблицата с данни за проекта (Фигура 9.3-2 от предишната глава) директно в LLM и да го помолим да построи прост модел за машинно обучение за нас.

- 👉 Изпратете текстова заявка до LLM чат (CHATGP, LlaMa, Mistral DeepSeek, Grok, Claude, QWEN:

Трябва да покажем изграждането на прост модел за машинно обучение за прогнозиране на разходите и времето за реализиране на нов проект X (фиг. 9.3-2 като приложено изображение) ↴

■ Отговор на LLM :



```

1  from sklearn.linear_model import LinearRegression
2  import numpy as np
3
4  # Data from the image
5  # Features: Number of apartments, number of floors, project complexity
6  X = np.array([
7      [5, 7, 2],
8      [8, 9, 6],
9      [3, 5, 3]
10 ])
11
12 # Targets: Time (days), Cost ($)
13 y_time = np.array([270, 330, 230])
14 y_cost = np.array([4502000, 5750000, 3262000])
15
16 # Creating linear regression models for time and cost
17 model_time = LinearRegression()
18 model_cost = LinearRegression()
19
20 # Training the models
21 model_time.fit(X, y_time)
22 model_cost.fit(X, y_cost)
23
24 # The new project's features
25 X_new = np.array([[4, 4, 7]])
26
27 # Predicting the time and cost for the new project
28 predicted_time = model_time.predict(X_new)
29 predicted_cost = model_cost.predict(X_new)
30
31 predicted_time, predicted_cost

```

Result  
`(array([238.44444444]), array([3042337.77777778]))`

**Фигура 9.3-4 LLM избра линейна регресия, за да създаде модел за машинно обучение за прогнозиране на разходите и времето за изпълнение на проекта.**

LLM автоматично разпознава таблицата от приложеното изображение и преобразува данните от визуален формат в табличен масив (фиг. 9.3-4 - ред 6). Този масив беше използван като основа за създаване на признания и етикети, от които беше създаден модел за машинно обучение (фиг. 9.3-4 - 17-ти-22-ри ред), който използва линейна регресия.

С помощта на базов модел на линейна регресия, обучен върху "изключително малък" набор от данни, бяха направени прогнози за нов хипотетичен строителен проект, обозначен като проект X. В

нашата задача този проект се характеризира с 40 апартамента, 4 етажа и ниво на сложност 7 (фигура 9.3-2).

Както се предвижда от линейния регресионен модел, базиран на ограничен и малък набор от данни за новия проект X (фигура 9.3-4 - линия 24-29):

- **Продължителността на строителството** ще бъде приблизително 238 дни (238.4444444)
- **Общият размер на разходите** ще бъде приблизително \$ 3,042,338 (3042337.777)

За по-нататъшно проучване на хипотезата за разходите по проекта е полезно да се експериментира с различни алгоритми и методи за машинно обучение. Ето защо нека прогнозираме същите стойности на разходите и времето за нов проект X въз основа на малък набор от исторически данни, като използваме алгоритъма K-Nearest Neighbours (k-NN).

## **Прогнозиране на разходите и времето за изпълнение на проекта с помощта на алгоритъма на най-близките съседи (k-NN)**

Използваме алгоритъма k-Nearest Neighbours (k-NN) като допълнителен предсказващ фактор, за да оценим разходите и продължителността на нов проект. Алгоритъмът K-Nearest Neighbours (k-NN) е метод за машинно обучение с наблюдение (supervised machine learning) както за класификация, така и за регресия. По-рано обсъждахме алгоритъма k-NN и в контекста на търсенето на векторни бази данни (фигура 8.2-2), където той се използва за намиране на най-близките вектори (например текстове, изображения или технически описание). При този подход всеки проект се представя като точка в многомерно пространство, където всяко измерение съответства на специфичен атрибут на проекта.

В нашия случай, като имаме предвид трите атрибути на всеки проект, ще ги представим като точки в триизмерно пространство (Фигура 9.3-5). Така нашият предстоящ проект X ще бъде локализиран в това пространство с координати ( $x=4, y=4, z=7$ ). Трябва да се отбележи, че в реални условия броят на точките и размерността на пространството могат да бъдат на порядъци по-големи.

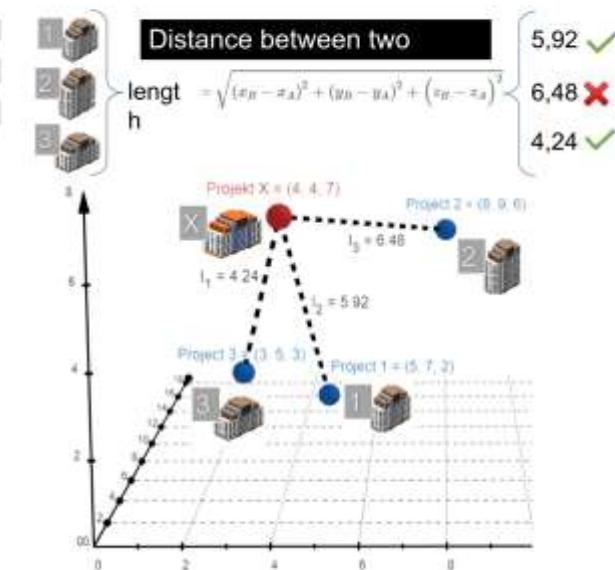
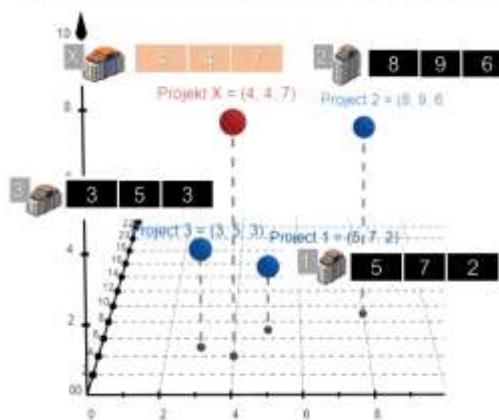
Алгоритъмът K-NN (k-най-близки съседи) работи чрез измерване на разстоянието между желания проект X и проектите в базата данни за обучение. Чрез сравняване на тези разстояния алгоритъмът определя проектите, които са най-близо до точката на новия проект X.

Например, ако вторият проект ( $x=8, y=9, z=6$ ) от нашата първоначална съвкупност от данни е много по-далеч от X (Фигура 9.3-5), отколкото другите проекти, той може да бъде изключен от по-нататъшния анализ. В резултат на това за изчисленията могат да се използват само двата ( $k=2$ ) най-близки проекти, въз основа на които ще се определи средната стойност.

Подобен метод, чрез търсене в съседство, позволява да се оценят приликите между проектите, което от своя страна помага да се направят заключения за вероятните разходи и срокове на нов проект въз основа на подобни проекти, които са били изпълнени преди това.

### k-nearest neighbors algorithm

The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.



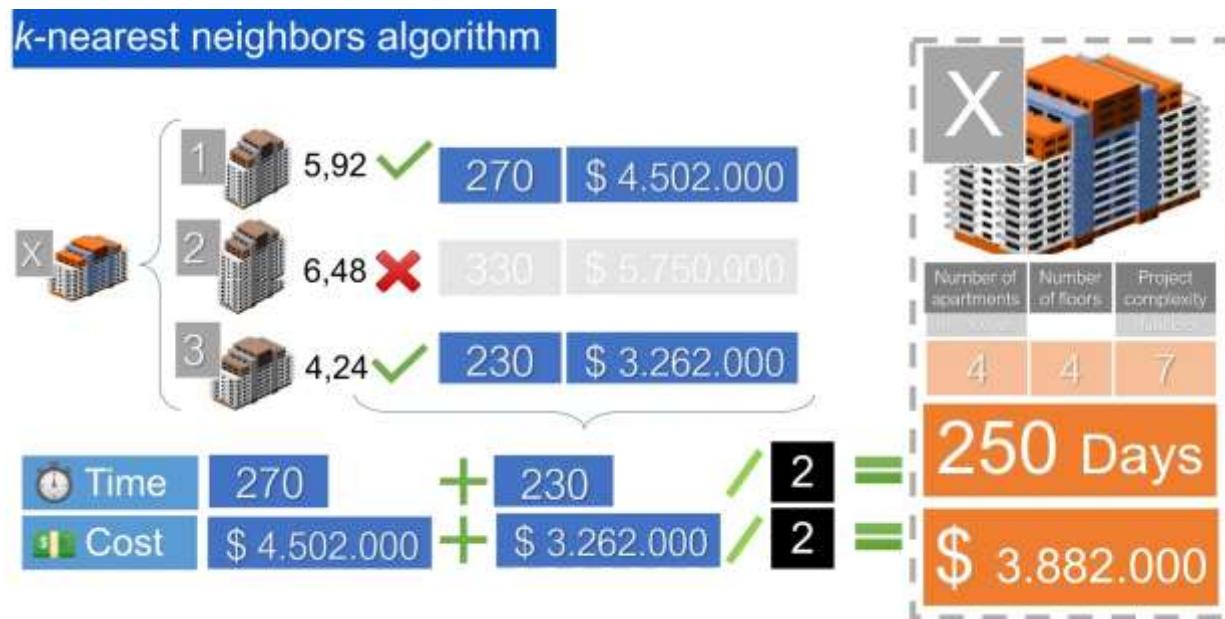
Фиг. 9.3-5 В алгоритъма K-NN проектите се представят като точки в многомерно пространство и най-близките проекти се избират въз основа на разстоянията за оценка на сходството и прогнозиране.

Работата на k-NN включва няколко ключови стъпки:

- **Подготовка на данните:** първо се зареждат набори от данни за обучение и тестване. Данните за обучение се използват за "трениране" на алгоритъма, а тестовите данни се използват за проверка на ефективността му.
- **Избор на параметъра K:** избира се числото K, което показва колко най-близки съседи (точки от данни) трябва да се вземат предвид в алгоритъма. Стойността на "K" е много важна, тъй като оказва влияние върху резултата.
- Процес на класификация и регресия за тестови данни:
  - **Изчисляване на разстоянията:** за всеки елемент от тестовите данни се изчислява разстоянието до всеки елемент от данните за обучение (Фигура 9.3-5). За тази цел могат да се използват различни методи за измерване на разстояния, като например Евклидово разстояние (най-често използваният метод), разстояние на Манхатън или разстояние на Хаминг.
  - **Сортиране и избор на K най-близки съседи:** след изчисляване на разстоянията те се сортират и се избират K най-близки точки до тестовата точка.
  - **Определяне на класа или стойността на тестова точка:** ако това е задача за класификация, класът на тестовата точка се определя въз основа на най-честия клас сред K избрани съседи. Ако това е задача за регресия, се изчислява средната стойност (или друга мярка за централна тенденция) на стойностите на K съседи.
- **Завършване на процеса:** след като всички тестови данни са класифицирани или за тях са направени прогнози, процесът е завършен.

Алгоритъмът k-nearest neighbours (k-NN) е ефективен в много практически приложения и е един от основните инструменти в арсенала на специалистите по машинно обучение. Този алгоритъм е популярен поради своята простота и ефективност, особено в задачи, при които връзките между данните са лесни за интерпретиране.

В нашия пример, след прилагане на алгоритъма на най-близките съседи K, бяха идентифицирани двата проекта (от нашата малка извадка) с най-късо разстояние до проекта X (Фигура 9.3-5). Въз основа на тези проекти алгоритъмът определя средната стойност на тяхната цена и продължителност на строителството. След анализ (фиг. 9.3-6) алгоритъмът, осреднявайки най-близките съседи, стига до заключението, че проектът X ще струва приблизително \$ 3 800 000 USD и ще отнеме приблизително 250 дни за изпълнение.

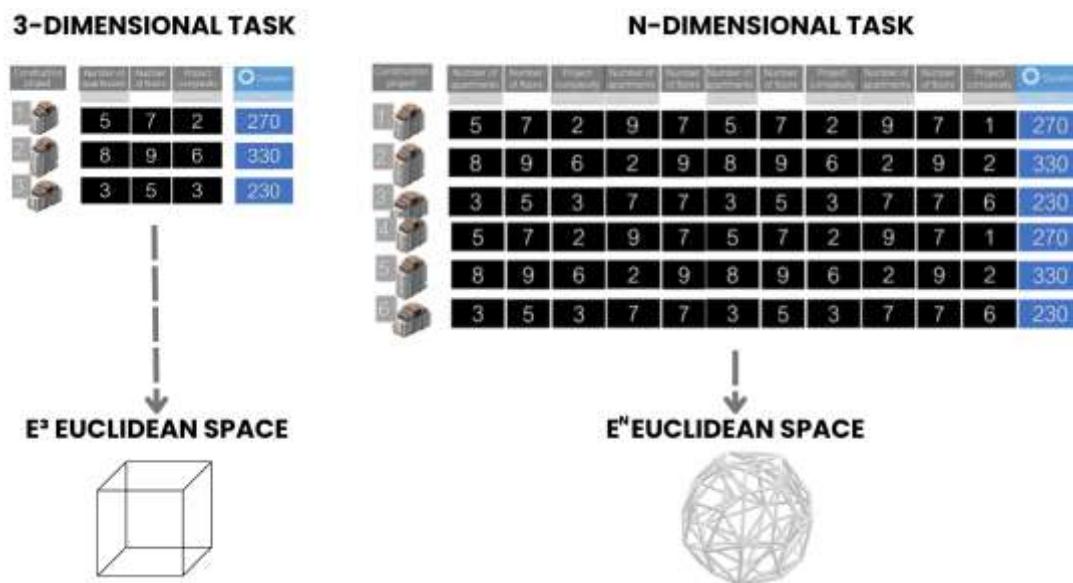


Фигура 9.3-6 Алгоритъмът К-блиズки съседи определя разходите и графика на проект X, като анализира двата най-близки проекта в извадката.

Алгоритъмът k-Nearest Neighbors (k-NN) е особено популярен в задачи за класификация и регресия, като например системи за препоръчване, където се използва за предлагане на продукти или съдържание въз основа на предпочтения, сходни с интересите на даден потребител. Освен това k-NN се използва широко в медицинската диагностика за класифициране на видове заболявания въз основа на симптомите на пациента, в разпознаването на образи и във финансовия сектор за оценка на кредитоспособността на клиентите.

Дори при ограничени данни моделите за машинно обучение могат да предоставят полезни прогнози и значително да подобрят аналитичния компонент на управлението на строителни проекти. С разширяването и изчистването на историческите данни е възможно да се премине към по-сложни модели - например отчитане на вида на строителството, местоположението, сезона на започване на строителството и други фактори.

Нашата опростена задача използва три атрибута за визуализация в 3D пространството, но реалните проекти средно включват стотици или хиляди атрибути (вж. набора от данни от главата "Пример за големи данни, базирани на CAD (BIM)", което значително увеличава размерността на пространството и сложността на представянето на проектите като вектори (фигура 9.3-7).



**Фигура 9.3-7 В опростения пример са използвани три атрибута за 3D -визуализация, докато в реалните проекти те са повече.**

Прилагането на различни алгоритми към същия набор от данни за проект X, който има 40 апартамента, 4 етажа и ниво на сложност 7, дава различни прогнозни стойности. Алгоритъмът за линейна регресия прогнозира време за завършване от 238 дни и разходи от \$ 3 042 338 (фигура 9.3-4), докато алгоритъмът k-NN прогнозира 250 дни и \$ 3 882 000 (фигура 9.3-6).

Точността на прогнозите, получени с помощта на модели за машинно обучение, зависи пряко от обема и качеството на входните данни. Колкото повече проекти са включени в обучението и колкото по-пълно и точно са представени техните характеристики (атрибути) и резултати (етикети), толкова по-голяма е вероятността за получаване на надеждни прогнози с минимални стойности на грешката.

Техниките за предварителна обработка на данни играят важна роля в този процес, като включват:

- Нормализация за привеждане на характеристиките към обща скала;

- Откриване и отстраняване на отклонения, премахване на изкривяването на модела;
- Кодиране на категорични атрибути, за да може да се манипулират текстови данни;
- Попълване на липсващи стойности, което повишава надеждността на модела.

Освен това се използват методи за кръстосано валидиране, за да се оцени обобщаването на модела и неговата устойчивост към нови набори от данни, за да се открие прекомерното приспособяване и да се подобри надеждността на прогнозата.

*Хаосът е ред, който трябва да бъде разшифрован [160].*

- Жозе Сарамаго, "Двойникът"

Дори да ви се струва, че хаосът на задачите ви не може да бъде описан формално, трябва да знаете, че всяко събитие в света и особено строителните процеси се подчиняват на математически закони, които може да се нуждаят от подкрепата на изчисляване на стойности не чрез строги формули, а с помощта на статистически данни и исторически данни.

Както традиционните оценки на отдела за изчисляване на разходите, така и моделите за машинно обучение неизбежно се сблъскват с несигурност и потенциални източници на грешки. Въпреки това, ако са налични достатъчно качествени данни, моделите за машинно обучение могат да демонстрират сравнима, а понякога дори по-висока точност на прогнозиране от експертните оценки.

Машинното обучение вероятно ще се превърне в надежден допълнителен инструмент за анализи, който може да: прецизира изчисленията, да предлага алтернативни сценарии и да идентифицира скрити зависимости между параметрите на проекта. Такива модели няма да претендират за универсалност, но скоро ще заемат важно място в изчисленията и процесите на вземане на решения. Технологиите за машинно обучение няма да изключат участието на инженерите, оценителите и анализаторите, а напротив - ще разширят техните възможности, като предложат допълнителна гледна точка, основана на исторически данни.

Ако бъде правилно интегрирано в бизнес процесите на строителните компании, машинното обучение има потенциала да се превърне във важен елемент в системата за подпомагане на вземането на управленски решения - не като заместител на хората, а като продължение на тяхната професионална интуиция и инженерна логика.

## Следващи стъпки: от съхранение до анализ и прогнозиране

Съвременните подходи за работа с данни започват да променят начина, по който се вземат решения в строителната индустрия. Преминаването от интуитивни оценки към обективен анализ на данните не само подобрява точността, но и открива нови възможности за оптимизиране на процесите. За да обобщим тази част, си струва да подчертаем основните практически стъпки, които ще ви помогнат да прилагате разгледаните методи в ежедневните си задачи:

■ Създаване на устойчива инфраструктура за съхранение

- Опитайте се да обедините различни документи и данни за проекта в единен табличен модел, като обедините ключовата информация в една рамка за данни за по-нататъшен анализ.
- Използване на ефективни формати за съхранение на данни - например колонарни формати като Apache Parquet вместо CSV или XLSX - особено за набори, които потенциално могат да се използват за обучение на модели за машинно обучение в бъдеще.
- Създаване на система за създаване на версии на данните за проследяване на промените по време на проекта

■ Внедряване на инструменти за анализ и автоматизация

- Започнете да анализирате исторически данни за проекти - по документация, модели, оценки - за да идентифицирате модели, тенденции и аномалии.
- Овладяване на процесите ETL (Extract, Transform, Load) за автоматично зареждане и подготовка на данните.
- Научете как да визуализирате ключови показатели с помощта на различни безплатни библиотеки за визуализация на Python
- Започнете да прилагате статистически методи и случаини извадки за получаване на представителни и възпроизвеждими аналитични резултати.

■ Повишаване на зрелостта при работа с данни

- Научете няколко основни алгоритъма за машинно обучение с прости и ясни примери, като например набора от данни за "Титаник"
- Анализирайте текущите процеси и определете къде можете да преминете от строга логика на причините и следствията към статистически методи за прогнозиране и оценяване.
- Започнете да третирате данните като стратегически актив, а не като страничен продукт: изграждайте процесите на вземане на решения на базата на модели на данни, а не на базата на конкретни софтуерни решения.

Строителните компании, които са осъзнали стойността на данните, навлизат в нова фаза на развитие, в която конкурентното предимство се определя не от количеството ресурси, а от скоростта на вземане на решения въз основа на анализи.



## МАКСИМАЛНО УДОБСТВО С ПЕЧАТНАТА ВЕРСИЯ

Вие притежавате безплатна цифрова версия на **Data-Driven Construction**. За по-удобна работа и бърз достъп до материалите ви препоръчваме да обърнете внимание на **печатното издание**:



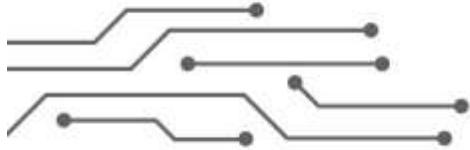
**■ Винаги под ръка:** книгата в печатен формат ще се превърне в надежден работен инструмент, който ще ви позволи бързо да намерите и използвате правилните визуализации и диаграми във всяка работна ситуация.

**■ Висококачествени илюстрации:** всички изображения и графики в печатното издание са представени с максимално качество.

**■ Бърз достъп до информация:** лесна навигация, възможност за правене на бележки, отметки и работа с книгата навсякъде.

Закупувайки пълната печатна версия на книгата, вие получавате удобен инструмент за комфортна и ефективна работа с информация: възможност за бързо използване на визуални материали в ежедневните задачи, бързо намиране на необходимите схеми и водене на бележки. Освен това с покупката си подкрепяте разпространението на отвореното знание.

Поръчайте печатна версия на книгата на адрес: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## Х ЧАСТ

# СТРОИТЕЛНАТА ИНДУСТРИЯ В ЦИФРОВАТА ЕРА. ВЪЗМОЖНОСТИ И ПРЕДИЗВИКАТЕЛСТВА

Последната десета част е цялостен поглед към бъдещето на строителната индустрия в ерата на цифровата трансформация. В нея се анализира преминаването от причинно-следствен анализ към работа с корелации на големи обеми данни. Правят се паралели между еволюцията на изобразителното изкуство и развитието на работата с данни в строителството, като се демонстрира как индустрията преминава от детайлен контрол към цялостно разбиране на процесите. Разглеждана е концепцията за "уберизация" на строителната индустрия, при която прозрачността на данните и автоматизираните изчисления могат радикално да променят традиционните бизнес модели, като премахнат необходимостта от посредници и намалят възможностите за спекуляции. Нерешените въпроси, като например универсалната класификация на елементите на, се обсъждат подробно, като се дава време на строителните компании да се адаптират към новата среда. Частта завършва с конкретни препоръки за оформяне на стратегия за цифрова трансформация, като включва анализ на уязвимите места и разширяване на услугите, за да се запази конкурентоспособността в променящата се индустрия.

## ГЛАВА 10.1.

### СТРАТЕГИИ ЗА ОЦЕЛЯВАНЕ: ИЗГРАЖДАНЕ НА КОНКУРЕНТНИ ПРЕДИМСТВА

#### Корелации вместо изчисления: бъдещето на строителните анализи

Поради бързата дигитализация на информацията (фигура 1.1-5) съвременното строителство претърпява фундаментална трансформация, при която данните се превръщат не просто в инструмент, а в стратегически актив, който може да промени из основи традиционните подходи към управлението на проекти и бизнес.

В продължение на хиляди години строителните дейности са се основавали на детерминистични методи - точни изчисления, детайли и строг контрол на параметрите. През първите векове от н.е. римските инженери прилагат математически принципи при изграждането на акведукти и мостове. През Средновековието архитектите се стремят към идеални пропорции на готическите катедрали, а в процеса на индустриализация през XX век се формират системи от стандартизираны норми и правила, които стават основа на масовото строителство

Днес векторът на развитието се измества от търсенето на строго причинно-следствени връзки към вероятностен анализ, търсене на корелации и скрити модели. Индустрията навлиза в нова фаза - данните се превръщат в ключов ресурс, а базирианият на тях анализ измества интуитивните и локално оптимизирани подходи.



**Фиг. 10.1-1 Скритият потенциал на данните за строителството: съществуващите в компанията изчисления са само върхът на айсберга, който ръководството може да анализира.**

Информационната система на една компания е като айсберг (фиг. 10.1-1): само малка част от потенциала на данните е видима за ръководството на компанията, докато основната стойност е скрита в дълбините. Важно е данните да се оценяват не само по отношение на текущото им използване, но и по отношение на възможностите, които ще разкрият в бъдеще. Именно тези компании, които се научат как да извличат скрити модели и да създават нови прозрения от данните, ще създадат устойчиво конкурентно предимство

Откриването на скрити модели и осмислянето на данни не е просто работа с числа, а творчески процес, който изисква абстрактно мислене и способност да се види цялата картина зад различни елементи. В този смисъл еволюцията на работата с данни може да се сравни с еволюцията на визуалните изкуства (фигура 10.1-2).

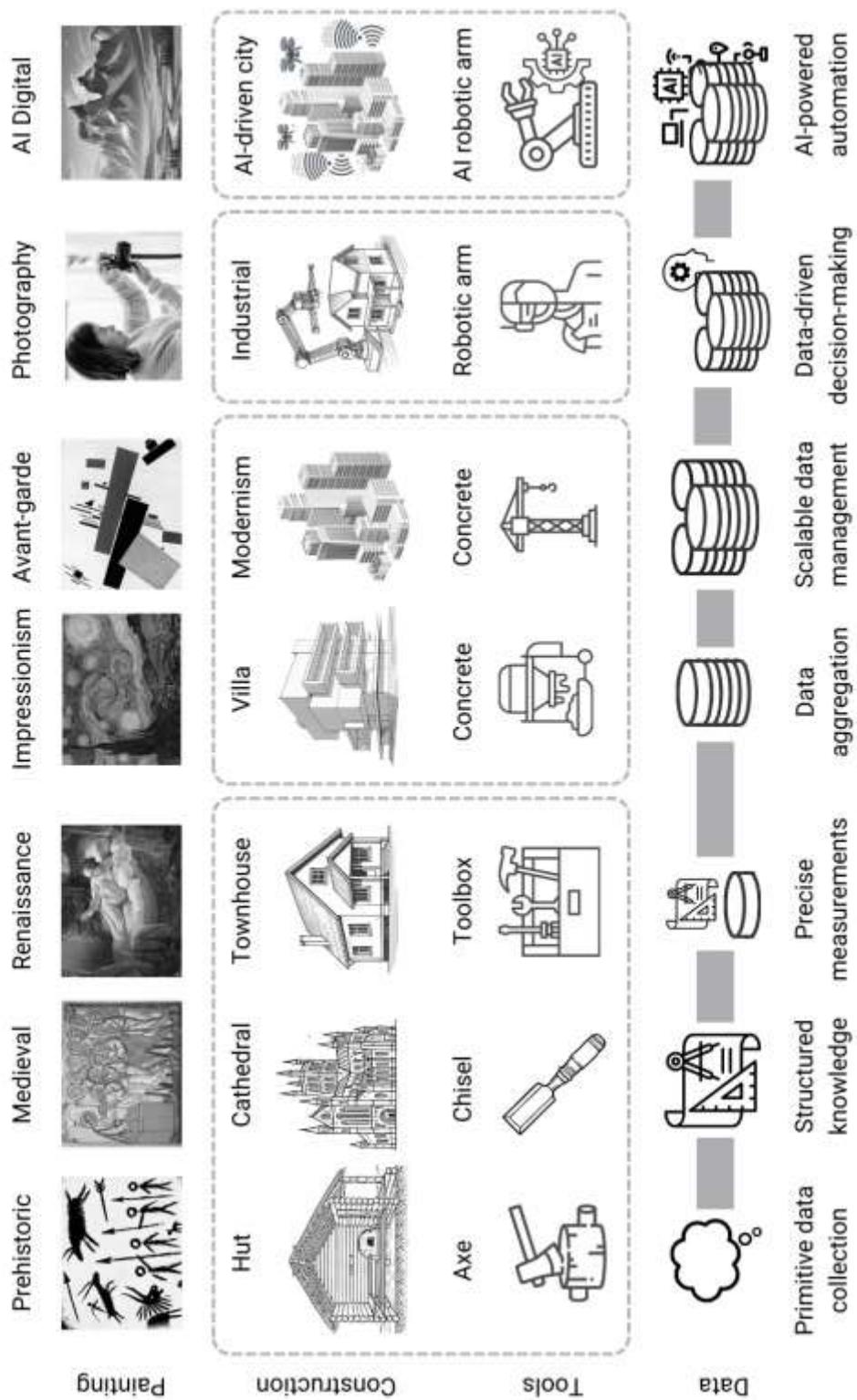
Развитието на строителството е много подобно на развитието на изобразителното изкуство. И в двата случая човечеството е напреднало от примитивни методи до сложни технологии за визуализация и анализ. В праисторически времена хората са използвали пещерни рисунки и примитивни инструменти, за да решават ежедневни задачи. През Средновековието и Ренесанса нивото на усъвършенстване на архитектурата и изкуството рязко се повишава. До началото на Средновековието строителните инструменти се развиват от обикновена брадва до обширни комплекти инструменти, символизиращи развитието на техническите познания.

Епохата на реализма е първата революция в изобразителното изкуство: художниците се научават

да възпроизвеждат и най-малките детайли, за да постигнат максимална достоверност. В строителството аналогът на този период са прецизните инженерни техники, подробните чертежи и строго регламентираните изчисления, които се превръщат в основа на проектантската практика в продължение на векове.

По-късно импресионизът променя самото възприемане на художествената реалност: вместо буквально да изобразяват формата, художниците започват да улавят настроението, светлината и динамиката, като се стремят да предадат цялостно впечатление, а не абсолютна точност. По подобен начин машинното обучение в областта на анализа на сгради преминава от строги логически модели към разпознаване на образи и вероятностни модели, които позволяват да се "видят" скрити зависимости в данните, недостъпни при класическия анализ. Този подход повтаря идеите на Bauhaus за минимализъм и функционалност, където смисълът (функцията) е по-важен от формата. Bauhaus се стреми да премахне излишното, да изостави орнаментиката в името на яснотата, утилитарността и масовата привлекателност. Нещата трябваше да са разбираеми и полезни, без излишъци - естетиката се раждаше от логиката на дизайна и предназначението.

С появата на фотографията в края на XIX в. изкуството се сдобива с нов инструмент за безпрецедентно точно заснемане на реалността и преобръща отношението към визуалните изкуства. По подобен начин в строителството индустрислната революция през XXI в. води до използването на роботизирани технологии, лазери, IoT, RFID и концепции като Connected Construction, при които събирането на отделни параметри еволюира до мащабирамо интелигентно заснемане на пълната реалност на строителната площадка.



Фиг. 10.1-2 Ерата на еволюциите в изобразителното изкуство е в съответствие с развитието на подходите за работа с данни в строителната индустрия.

Днес, както визуалните изкуства преживяват преосмисляне с появата на инструментите за изкуствен интелект и LLM, така и строителната индустрия преживява друг квантов скок: интелигентните системи, управляеми от изкуствен интелект (AI), LLM чат позволяват прогнозиране, оптимизиране и генериране на решения с минимална човешка намеса.

Ролята на данните в проектирането и управлението се е променила коренно. Докато преди знанията се предаваха устно и имаха емпиричен характер - точно както реалността е била улавяна чрез ръчно рисувани картини до XIX век - днес фокусът е върху пълното цифрово улавяне на строителната "картина". С помощта на алгоритми за машинно обучение тази цифрова картина се превръща в импресионистично представяне на строителната реалност - не точно копие, а обобщено, вероятностно разбиране на процесите.

Бързо се приближаваме към епоха, в която процесите на проектиране, изграждане и експлоатация на сгради не просто ще бъдат разширени, а до голяма степен ще се управяват от системи с изкуствен интелект. Точно както съвременното дигитално изкуство се създава без четка - с помощта на текстови подсказки и генеративни модели - архитектурните и инженерните решения на бъдещето ще се оформят от ключови запитвания и параметри, зададени от потребителя.

През 21-ви век достъпът до данни, тяхното тълкуване и качеството на анализите стават задължителни за успеха на проектите. А стойността на данните се определя не от техния обем, а от способността на специалистите да ги анализират, проверяват и превръщат в действия.

## **Подход, основан на данни, в строителството: ново ниво на инфраструктурата**

В историята на човечеството всеки подобен технологичен скок е довел до фундаментални промени в икономиката и обществото. Днес сме свидетели на нова вълна на трансформация, сравнима по мащаби с индустриалната революция от XIX век. Но докато преди сто години основният двигател на промяната са били механичните сили и енергийните технологии, сега това са данните и изкуственият интелект.

Агентите за машинно обучение, LLM и AI променят самата същност на приложенията, като правят ненужни традиционните софтуерни стекове (разгледани във втората част на книгата) (фиг. 2.2-3). Цялата логика на данните е съсредоточена в агентите на ИИ, а не в твърдо кодирани бизнес правила (фиг. 2.2-4).

В епохата на данните традиционните възгледи за приложенията се променят из основи. Преминаваме към модел, при който обемистите, модулни корпоративни системи неизбежно ще отстъпят място на отворени, леки и персонализирани решения.

В бъдеще ще остане само основната структура от данни, а цялото взаимодействие с нея ще се осъществява чрез агенти, работещи директно с базата данни. Наистина вявам, че целият стек от приложения ще изчезне, защото просто няма да има нужда от него, когато изкуственият интелект взаимодейства директно с основната база данни. Цялата си кариера съм прекарал в SaaS - изграждах компании, работех в тях и ако трябва да бъда честен, вероятно сега не бих стартиран нов SaaS бизнес. И вероятно не бих инвестирил в компании за SaaS точно сега. Ситуацията е твърде несигурна. Това не означава, че в бъдеще няма да има софтуерни компании, просто те ще изглеждат по съвсем различен начин. Бъдещите системи ще бъдат бази данни с бизнес логика, въведена в [AI] агенти. Тези агенти ще работят с множество хранилища на данни едновременно, като няма да се ограничават до една единствена база данни. Цялата логика ще се премести в слоя на изкуствения интелект [46].

- Матю Берман, главен изпълнителен директор на Forward Future

Основната разлика в новата парадигма е свеждането до минимум на технологичния баласт. Вместо монументални сложни и затворени софтуерни системи ще получим гъвкави, отворени и бързо адаптируеми модули, които буквально "живеят" в потока от данни (фиг. 7.4-1 - Apache Airflow, NiFi). Архитектурата на бъдещото управление на процесите предвижда използването на микроприложения - компактни, специално създадени инструменти, коренно различни от масивните и затворени ERP, PMIS, CDE, CAFM системи. Новите агенти ще бъдат възможно най-адаптивни, интегрирани и специфични за бизнеса (например Low-Code/No-Code Фиг. 7.4-6).

Цялата бизнес логика ще бъде предадена на тези агенти [AI] и те ще извършват CRUD [Create, Read, Update, and Delete] операции върху множество хранилища, което означава, че няма да правят разлика коя бекенд система се използва. Те ще актуализират множество бази данни и цялата логика ще се окаже на така нареченото AI -ниво. И след като нивото на изкуствения интелект се превърне в място, където се намира цялата логика, хората ще започнат да заменят бекендите. Вече виждаме доста висок процент на пазарни победи в използването на бекендове и агенти на Dynamics и ще се движим агресивно в тази посока, опитвайки се да обединим всичко това. Независимо дали става дума за обслужване на клиенти или за други области, например не само CRM, но и за нашите финансови и оперативни решения. Защото хората искат повече бизнес приложения, управлявани от ИИ, при които логическият слой може да се управлява от ИИ и агенти с ИИ. [...]. Едно от най-вълнуващите неща за мен е Excel с Python, което е сравнено с GitHub с Copilot. Тоест това, което направихме: сега, когато имате Excel, трябва просто да го отворите, да стартирате Copilot и да започнете да си играете с него. Вече не става дума само за разбиране на наличните числа - той сам ще направи план. Точно както работното пространство на GitHub Copilot създава план и след това го изпълнява, то е като анализатор на данни, който използва Excel като инструмент за визуализиране на редове и колони за анализ. Така че Copilot използва Excel като инструмент с всичките му възможности, защото може да генерира данни и има интерпретатор на Python.

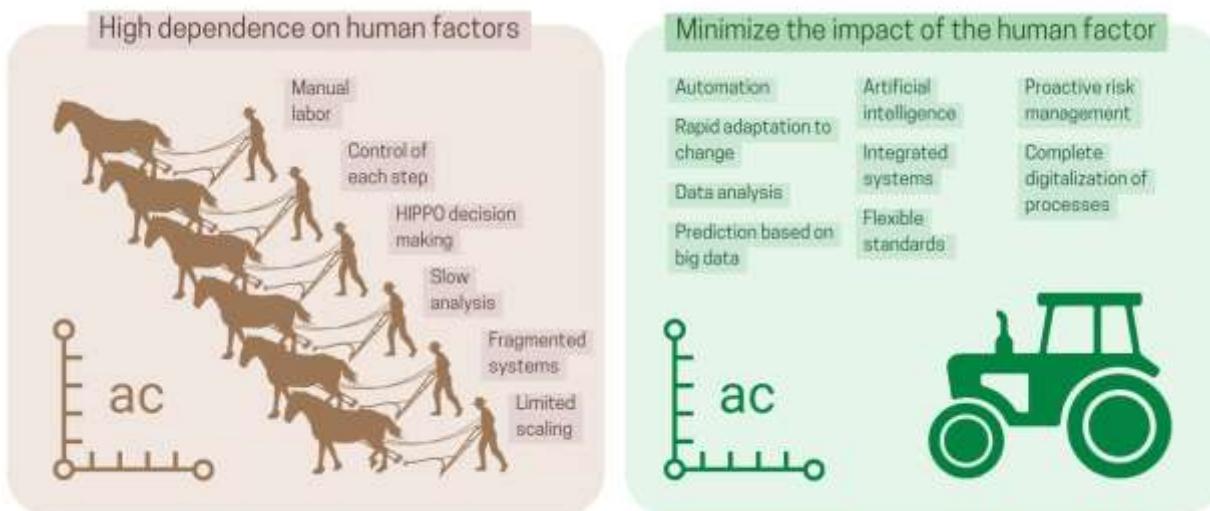
- Сатя Надела, главен изпълнителен директор, Microsoft, интервю за канал BG2, декември 2024 г. [28]

Трансформацията, на която сме свидетели в логиката на офис приложенията - преминаване от модулни, затворени системи към агенти с изкуствен интелект, работещи директно с отворени данни - е само част от много по-голям процес. Не става въпрос само за промяна на интерфейсите или софтуерната архитектура: промените ще засегнат основните принципи на организацията на работата, вземането на решения и управлението на бизнеса. В строителството това ще доведе до логика, основана на данни, в която данните стават централна част от процесите от проектирането до управлението на ресурсите и строителния мониторинг.

## **Цифровият офис от следващо поколение: как AI променя работното пространство**

Преди почти век човечеството вече преживява подобна технологична революция. Преходът от парни машини към електрически двигатели отне повече от четири десетилетия, но в крайна сметка катализира безprecedентен ръст на производителността - най-вече чрез децентрализация на енергийния капацитет и гъвкавостта на новите решения. Тази промяна не само промени хода на историята, като премести по-голямата част от населението от селските към градските райони, но и постави основите на съвременната икономика. Историята на технологиите е път от физическия труд към автоматизацията и интелигентните системи. Точно както тракторът замени десетки култиватори на почвата, съвременните цифрови технологии заменят традиционните офис методи за управление на строителството (фигура 10.1-3). Още в началото на XX в. по-голямата част от населението на света обработва земята ръчно, докато през 30-те години на XX в. започва механизирането на

труда с машини и трактори.



**Фиг. 10.1-3** Както тракторът замени десетки хора в началото на 20-ти век, така и машинното обучение ще замени традиционните методи за управление на бизнеса и проектите през 21-ви век.

Точно както преди сто години човечеството е преминало от обработване на отделни парцели земя с примитивни инструменти към мащабно земеделие с помощта на машини, така и днес преминаваме от обработване на отделни "силози" с информация към работа с масиви от данни с помощта на мощнни "трактори" - ETL-линия и алгоритми с изкуствен интелект.

Намираме се на прага на подобен скок, но в цифровата сфера: от традиционно, ръчно управление на бизнеса към модели, базирани на данни.

Пътят към пълноценна архитектура, базирана на данни, ще изисква време, инвестиции и организационни усилия. Но този път открива пътя не просто към постепенно подобрение, а към качествен скок към по-голяма ефективност, прозрачност и управляемост на строителните процеси. Всичко това зависи от системното въвеждане на цифрови инструменти и отпадането на оstarели бизнес практики.

Параметризацията на задачите, ETL, LLM, компонентите на IoT, RFID, токенизацията, големите **данни** и машинното обучение ще превърнат традиционното строителство в **управлявано от данни строителство**, при което всеки детайл от проекта и строителния бизнес ще се контролира и оптимизира с помощта на данни.

Анализирането на информацията преди отнемаше хиляди човекочасове. Сега тези задачи се изпълняват от алгоритми и LLM, които превръщат разнородни масиви от данни в стратегически източници с помощта на подсказки. В света на технологиите се случва същото, което се случи със селското стопанство: преминаваме от мотика към автоматизиран агробизнес. Така и офисната работа в строителството - от файлове в Excel и ръчно обобщаване - преминава към интелигентна система, в която данните се събират, почистват, структурират и превръщат в прозрения.

Още днес компаниите трябва да започнат да "култивират" информационните полета чрез качествено събиране на данни и структуриране на информацията, да ги "наторяват" с инструменти за почистване и нормализиране, а след това да ги "прибират" под формата на прогнозни анализи и автоматизирани решения. Ако един съвременен фермер с машина е в състояние да замени стотина обработвачи на почвата, то интелигентните алгоритми ще могат да премахнат рутината от служителите и да ги прехвърлят в ролята на стратегически мениджъри на информационните потоци.

Важно е обаче да се разбере, че създаването на истинска организация, основана на данни, не е бърз процес. Това е дългосрочна стратегическа насока, подобна на създаването на нов терен за отглеждане на нова гора (фиг. 1.2-5) от системи, където всяко "дърво" в тази екосистема е отделен процес, компетентност или инструмент, който изисква време за израстване и развитие. Както и в случая с истинската гора, успехът зависи не само от качеството на посадъчния материал (технологията), но и от почвата (корпоративната култура), климата (бизнес средата) и грижите (системния подход).

Компаниите вече няма да могат да разчитат единствено на затворени, готови решения. За разлика от предишни етапи на технологично развитие, сегашният преход - към отворен достъп до данни, изкуствен интелект и отворен код - едва ли ще бъде подкрепен от големите доставчици, тъй като пряко застрашава техните утвърдени бизнес модели и основни източници на приходи.

Както показва проучването на Harvard Business School [40], което вече беше разгледано в главата за четвъртата и петата технологична революция, разходите за създаване на най-използваните решения с отворен код от нулата за всички компании биха били около 4,15 милиарда долара. Ако си представим обаче, че всяка компания ще разработи свои собствени алтернативи без достъп до съществуващите инструменти с отворен код, което се случва през последните десетилетия, общите разходи на бизнеса биха могли да достигнат колосалните 8,8 трилиона долара - това е цената на ирационалното търсене, на която може да бъде оценен софтуерният пазар.

Технологичният напредък неизбежно ще доведе до преосмисляне на установените бизнес модели. Докато преди компаниите можеха да печелят пари от сложни, непрозрачни процеси и затворени данни, с развитието на изкуствения интелект и анализите този подход става все по-малко жизнеспособен.

В резултат на демократизацията на достъпа до данни и инструменти традиционният пазар за продажба на софтуер може да се свие значително. В същото време обаче ще се разрасне нов пазар - пазарът на цифрова експертиза, персонализация, интеграция и проектиране на решения. Тук стойността няма да идва от продажбата на лицензи, а от способността да се изграждат гъвкави, отворени и адаптивни цифрови процеси. Както електрификацията и появата на тракторите породиха нови индустрии, така и прилагането на големи обеми от данни, изкуствен интелект и LLM ще открие напълно нови хоризонти за бизнеса в строителната индустрия, което ще изисква не само технологични инвестиции, но и дълбока трансформация на мисленето, процесите и организационните структури. И тези компании и специалисти, които осъзнават това и започнат да действат днес, ще бъдат лидерите на утешния ден.

В свят, в който отворените данни се превръщат в основен актив, наличието на информация ще промени правилата на играта. Инвеститорите, клиентите и регуляторните органи все повече ще изискват прозрачност, а алгоритмите за машинно обучение ще могат автоматично да идентифицират

несъответствия в оценките, сроковете и разходите. Това създава условия за нов етап на цифровата трансформация, който постепенно ни води към "уберизация" на строителния бранш.

## Отворените данни и Uberisation са заплаха за съществуващия строителен бизнес

Строителството се превръща в процес на управление на информацията. Колкото по-точни, качествени и пълни са данните, толкова по-ефективни са проектирането, изчисленията, оценките на разходите, монтажът и експлоатацията на сградите. В бъдеще ключовият ресурс няма да бъдат кранът, бетонът и арматурата, а способността да се събира, анализира и използва информация.

В бъдеще клиентите на строителните компании - инвеститори и клиенти, финансиращи строителството - неминуемо ще използват стойността на отворените данни и анализите на исторически данни. Това ще открие възможности за автоматизиране на изчисляването на сроковете и разходите по проектите, без да се ангажират строителните компании с въпроси, свързани с изчисляването на разходите, което ще помогне за побърз контрол на разходите и идентифициране на излишните разходи.

Представете си строителна площадка, на която лазерни скенери, квадрокоптери и фотограметрични системи събират точни данни в реално време за обема на използвания бетон. Тази информация автоматично се преобразува в прости плоски MESH -модели с метаданни, заобикаляйки тромавите CAD (BIM) системи, без да зависи от сложни геометрични ядра, ERP или PMIS. Тези данни, събрани от строителната площадка, се прехвърлят централизирано в единно структурирано хранилище, достъпно за клиента за независим анализ, където се качват реални цени от различни строителни цехове, с параметри, вариращи от процента на кредитно финансиране до динамично променящи се фактори като климатични условия, борсови котировки на строителни материали, логистични тарифи и статистически сезонни колебания в цените на труда. При тези условия всички несъответствия между проектните и действителните обеми на материалите стават незабавно очевидни, което прави невъзможно манипулирането на оценките както на етапа на проектиране, така и при предаването на проекта. В резултат на това прозрачността на строителния процес се постига не чрез армия от надзорници и мениджъри, а чрез обективни цифрови данни, които ще сведат до минимум човешкия фактор и възможността за спекулации.

В бъдеще този вид работа по контрола на данните ще се извършва по-скоро от мениджъри на данни от страна на клиента (фиг. 1.2-4 CQMS мениджър). Това важи особено за калкулациите и оценките на проектите: там, където преди е имало цял отдел от оценители, утре вече ще има инструменти за машинно обучение и прогнозиране, които ще определят ценовите граници, в които строителните компании да се вместят.

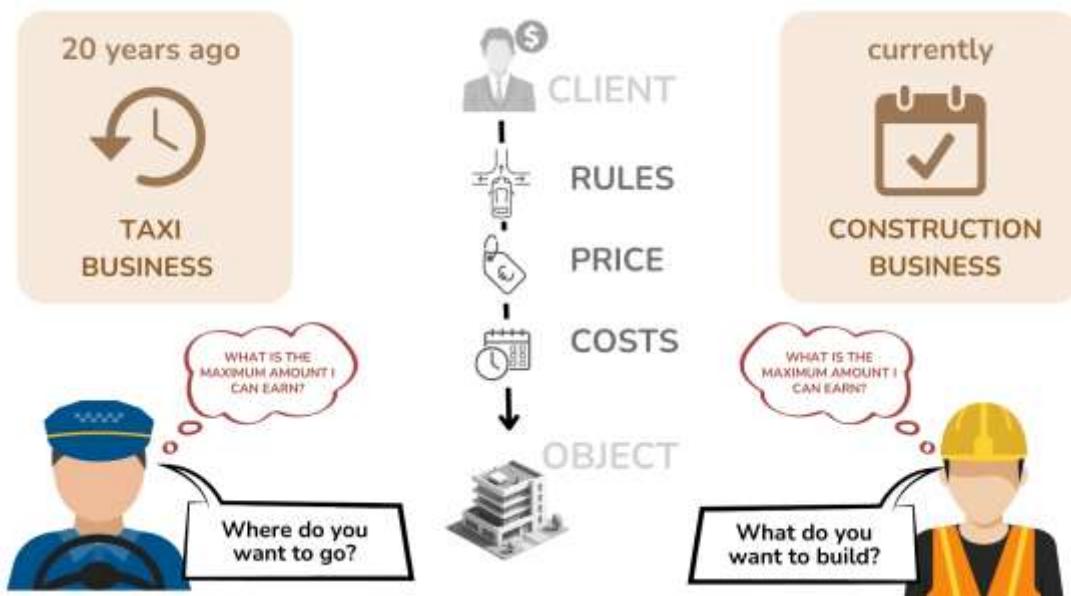
*Като се има предвид фрагментираният характер на [строителната] индустрия, в която повечето системи и подсистеми се доставят от МСП, цифровата стратегия трябва да идва от клиента. Клиентите трябва да създадат условия и механизми за отключване на цифровите възможности на веригата на доставки [20].*

- Андрю Дейвис и Джулиано Деникол, Accenture "Създаване на повече стойност чрез капиталови проекти"

Подобна откритост и прозрачност на данните представлява заплаха за строителните компании, които са свикнали да печелят пари от непрозрачни процеси и объркващи отчети, където спекулациите и скритите разходи могат да бъдат скрити зад сложни и затворени формати и модулни платформи за данни. Ето защо строителните компании, както и доставчиците, които популяризират решения с отворен код, едва ли ще се интересуват от пълното внедряване на отворените данни в своите бизнес процеси. Ако данните са достъпни и лесни за обработка от клиента, те могат да бъдат проверявани автоматично, което ще елиминира възможността за надценяване на обемите и манипулиране на оценките.

Според доклада на Световния икономически форум "Формиране на бъдещето на строителството" (2016 г.) [5], едно от основните предизвикателства пред отрасъла остава пасивната роля на клиента. Въпреки това именно клиентите трябва да поемат по-голяма отговорност за резултатите от проектите - от ранното планиране, през избора на устойчиви модели на взаимодействие, до мониторинга на изпълнението. Без активното участие на собствениците на проекти системната трансформация на строителната индустрия е невъзможна.

Загубата на контрол върху изчисленията на обема и разходите вече трансформира други индустрии през последните 20 години, позволявайки на клиентите директно, без посредници, да спъват своите цели. Дигитализацията и прозрачността на данните трансформираха много традиционни бизнес модели, като например таксиметровите шофьори с появата на Uber (фигура 10.1-4), хотелиерите с появата на Airbnb и търговците на дребно и магазини с появата на Amazon, както и банките с появата на необанките и децентрализираните финтех екосистеми, където директният достъп до информация и автоматизацията на изчисленията на обема и разходите значително намалиха ролята на посредниците.



Фиг. 10.1-4 Строителният бизнес ще се сблъска с уберизацията, с която се сблъскаха таксиметровите шофьори, хотелиерите и търговците на дребно преди 10 години.

Процесът на демократизиране на достъпа до данни и инструменти за тяхната обработка е неизбежен и с течение на времето отворените данни за всички компоненти на проекта ще се превърнат в изискване на клиентите и нов стандарт. Ето защо въпросите за въвеждане на отворени формати и прозрачни изчисления ще бъдат насярчавани от инвеститорите, клиентите, банките и фондовете за дялово участие (private equity) - тези, които са крайни потребители на изградените обекти и след това експлоатират обекта в продължение на десетилетия.

Големите инвеститори, клиенти и банки вече изискват прозрачност в строителния сектор. Според проучването на Accenture "Създаване на повече стойност чрез капиталови проекти" (2020 г.) [20] прозрачните и надеждни данни се превръщат в решаващ фактор за инвестиционните решения в строителството. Както посочват експертите, надеждното и ефективно управление на проекти е невъзможно без прозрачност, особено по време на кризи. Освен това собствениците на активи и изпълнителите все повече се насочват към договори, които стимулират споделянето на данни и съвместния анализ, отразявайки нарастващите изисквания от страна на инвеститорите, банките и регулаторните органи за отчетност и прозрачност.

Движението на инвеститора, на клиента от идеята до завършената сграда в бъдеще ще прилича на пътуване на автопилот - без шофър в лицето на строителна компания, обещава да стане независимо от спекулациите и несигурността.

Ерата на отворените данни и автоматизацията неминуемо ще промени строителния бизнес, както вече го направи в банковото дело, търговията, селското стопанство и логистиката. В тези отрасли ролята на посредниците и традиционните начини на правене на бизнес отстъпват място на автоматизацията и роботизацията, като не оставят място за неоправдани надценки и спекулации.

Данните и процесите във всички икономически дейности на човека не се различават от тези, с които трябва да се справят специалистите в строителния бранш. В дългосрочен план строителните компании, които днес доминират на пазара, като определят стандартите за цена и качество на услугите, могат да загубят ролята си на ключов посредник между клиента и неговия строителен проект.

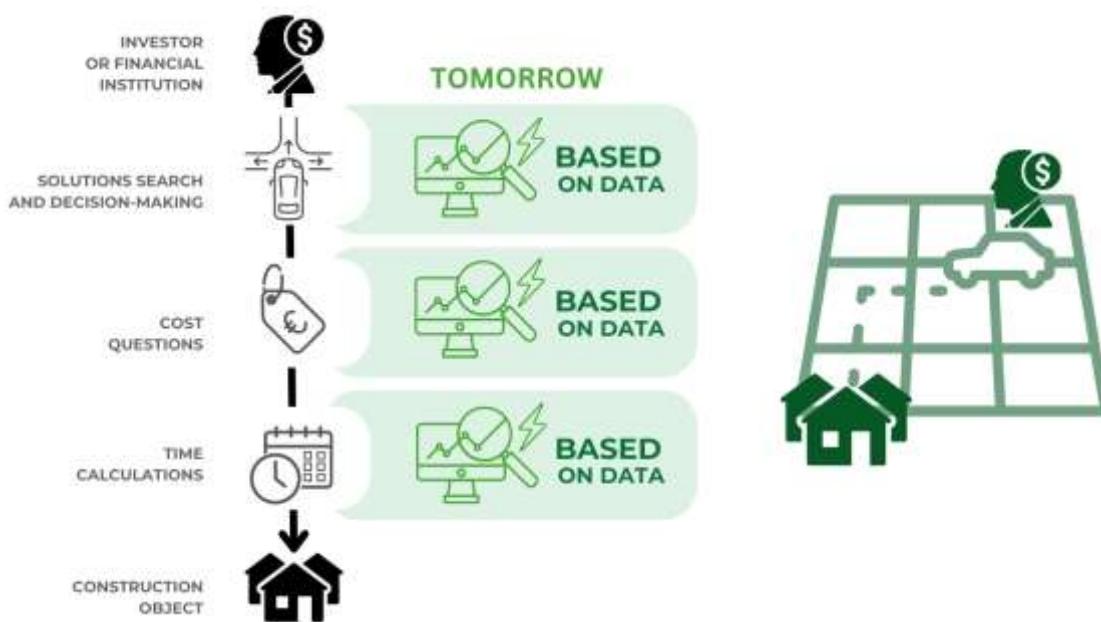
## **Неразрешените проблеми на уберализацията като последен шанс да се използва времето за трансформация**

Но нека се върнем към реалностите в строителния бранш. Докато в някои сектори на икономиката се появяват самоуправляващи се автомобили, децентрализирани финансови системи и решения, базирани на изкуствен интелект, значителна част от строителните компании все още остават организации, базирани на хартиени носители, в които ключовите решения се вземат по-скоро въз основа на интуицията и опита на отделните специалисти.

В тази парадигма съвременната строителна компания може да бъде сравнена с 20-годишна таксиметрова компания, която контролира ресурсите, маршрутите и времето за доставка и отговаря за времето и разходите на "пътуването" - от идеята за проекта (логистика и процес на монтаж) до доставката на проекта. Точно както GPS (в строителството IoT, RFID) и алгоритмите за машинно обучение в изчисленията на времето/разходите никога трансформираха областта на транспорта, данните, алгоритмите и AI-агентите имат потенциала да трансформират управлението на строителството - от интуитивни оценки до прогнозни, направлявани модели. През последните 20 години много отрасли - финанси, селско стопанство, търговия на дребно и логистика - постепенно елиминираха възможността за спекулации поради непрозрачността на данните. Цените, разходите за доставка или финансовите транзакции се изчисляват автоматично и статистически обосновано - само за няколко секунди на цифрови платформи.

С поглед към бъдещето строителните компании трябва да признаят, че демократизирането на достъпа до данни и инструментите за техния анализ ще наруши традиционния подход за оценка на разходите и сроковете на проектите и ще премахне възможността за спекулиране с непрозрачни данни за обема и цените.

Подобно на шофирането по регулиран път без намесата на шофьора, строителните процеси на бъдещето все повече ще приличат на "Uber" система - с автоматизирана оценка на времето и разходите, прозрачно насочване на задачите и минимална зависимост от човешкия фактор. Това ще промени самото естество на "пътуването" от идеята до реализацията - ще го направи по-предсказуемо, управляемо и основано на данни.



**Фиг. 10.1-5** Разходите и времето за "пътуване" по време на строителството ще се определят с помощта на машинно обучение и статистически инструменти.

С постепенното въвеждане на нови разпоредби и изисквания в почти всички страни по света, които задължават предаването на CAD- (BIM)-модели на клиенти или банки, финансиращи строителни проекти, клиентът и възложителят имат възможност самостоятелно да осигурят прозрачност на изчисленията на разходите и обхвата. Това е от особено значение за големите клиенти и инвеститори, които разполагат с достатъчно компетенции и инструменти за бърз анализ на обемите и наблюдение на пазарните цени. За компаниите, които изпълняват мащабни стандартни проекти - магазини, офис сгради, жилищни комплекси - подобни практики се превръщат в стандарт.

Тъй като информационното съдържание на моделите става все по-пълно и по-стандартизирано, възможността за манипулация и спекулация на практика изчезва. Цифровата трансформация постепенно променя правилата на играта в строителния бранш и компаниите, които не се адаптират към тези промени, могат да бъдат изправени пред сериозни предизвикателства.

Засилената конкуренция, технологичните сътресения и намаляването на маржовете могат да окажат влияние върху устойчивостта на бизнеса. При ограничена ликвидност все повече участници в индустрията се обръщат към автоматизацията, анализа и технологиите за данни като начин за подобряване на ефективността и прозрачността на процесите. Тези инструменти се превръщат във важен ресурс за запазване на конкурентоспособността в променящата се икономическа среда.

Може би не трябва да чакаме външните обстоятелства да наложат спешни действия - много по-ефективно е да започнем да се подготвяме още днес, като укрепваме цифровите компетенции, внедряваме модерни решения и изграждаме култура, ориентирана към данните.

Една от последните ключови технологични пречки пред мащабната цифрова трансформация на строителната индустрия, която ще засегне всяка компания през следващите години, е проблемът с автоматичната класификация на елементите на строителните проекти.

Без надеждна, точна и мащабируема класификация е невъзможно да се създаде основа за пълноценни анализи, автоматизация на процесите и управление на жизнения цикъл с помощта на ИИ и прогнозни модели. Докато класификацията на обектите все още зависи от ръчното им интерпретиране от опитни специалисти - майстори, проектанти, оценители - строителната индустрия все още разполага с прозорец от възможности. Това време може да се използва за подготовка за неизбежните промени: нарастващите изисквания за прозрачност, демократизацията на инструментите и данните и появата на системи за автоматична класификация, които радикално ще променят правилата на играта.

Задачата за автоматична класификация на елементите на строителния свят е сравнима по своята сложност с разпознаването на обекти в безпилотните системи за управление, което е едно от основните предизвикателства. Нека си представим безпилотен автомобил, който пътува от точка А до точка Б (фиг. 10.1-5). Настоящите системи за автоматично шофиране са затрупани от проблема за класифициране на обектите, които се разпознават от лидерите и камерите. Не е достатъчно автомобилът просто да "види" препятствие или ориентир, а трябва да може да разпознае дали това е пешеходец, пътен знак или кошче за боклук.

Подобно фундаментално предизвикателство стои и пред цялата строителна индустрия. Елементите на проекта - като прозорци, врати или колони - могат да бъдат записани в документацията, представени в CAD модели, заснети на строителната площадка или разпознати в облаци от точки от лазерно сканиране. Въпреки това тяхното визуално или грубо геометрично разпознаване не е достатъчно за изграждането на наистина автоматизирана система за управление на проекти. Необходимо е да се гарантира, че всеки елемент е точно и последователно категоризиран във вид, който е еднозначно разпознаваем във всички последващи процеси - от оценките и спецификациите до логистиката, контрола на запасите и най-важното - операциите (фиг. 4.2-6).

Именно на този етап - преходът от разпознаване към смислена класификация - възниква една от основните пречки. Дори ако цифровите системи са технически способни да различават и идентифицират обекти в моделите и на строителната площадка, основната трудност се състои в правилното и контекстуално стабилно определяне на типа на елемента за различните софтуерни среди.. Например една врата може да бъде обозначена от проектанта в CAD модел като елемент от категорията "врата", но при прехвърлянето ѝ в ERP или PMIS система може да бъде неправилно определена, било поради грешка на проектанта, било поради несъответствия между системите. Освен това елементът често губи някои важни атрибути или изчезва от счетоводството на системата по време на експорта и импорта на данни. Това води до пропуски в потока от данни и подкопава принципа за цялостна цифровизация на строителните процеси. Това създава критична разлика между "видимото" и "разбирамото" семантично значение, което подкопава целостта на данните и значително усложнява автоматизацията на процесите през целия жизнен цикъл на строителния проект.

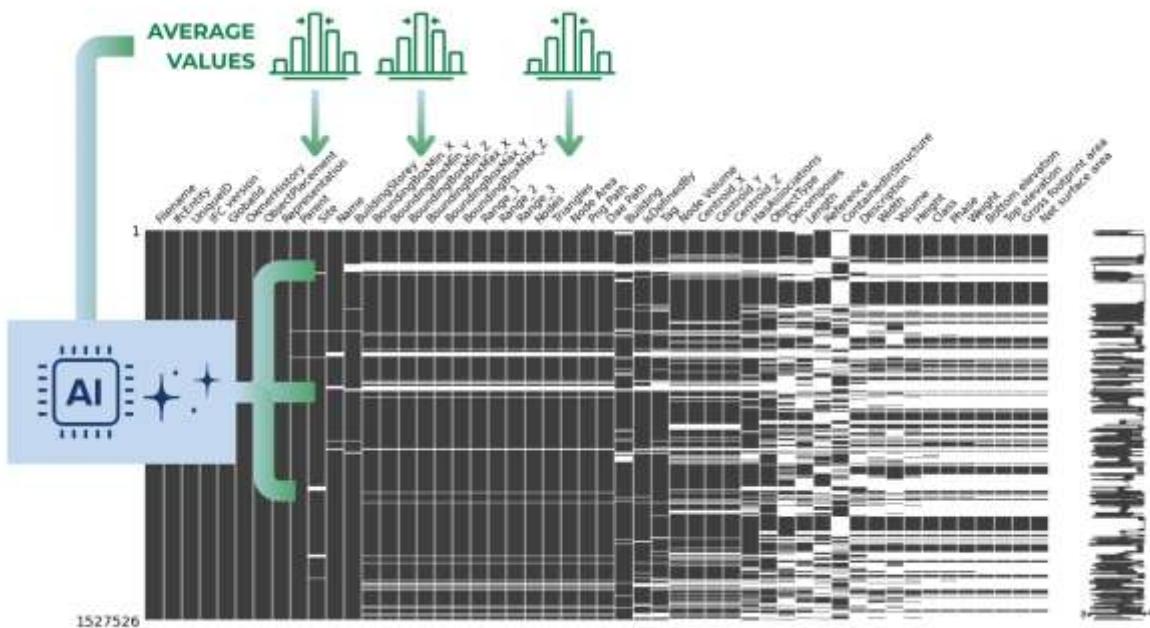
Решението на предизвикателството за универсална класификация на строителните елементи с помощта на технологии за големи обеми от данни и машинно обучение (фиг. 10.1-6) ще бъде катализатор за трансформация в цялата индустрия - и може би неочаквано откритие за много строителни компании. Единната, обучителна система за класификация ще бъде основата за мащабируеми ана-

лизи, цифрово управление и внедряване на изкуствения интелект в ежедневните практики на строителните организации.

NVIDIA и други технологични лидери вече предлагат решения в други индустрии, които могат автоматично да категоризират и структурират огромни количества текстова и визуална информация.

Моделът NeMo Curator на NVIDIA [161] например е специализиран в автоматичното класифициране и категоризиране на данни в предварително определени категории, като играе ключова роля в оптимизирането на конвейерите за обработка на информация за фина настройка и предварително обучение на генеративни модели на ИИ. Платформата Cosmos се обучава върху реални видео и 3D сцени [162], като осигурява основа за автономни системи и цифрови близнази, които вече се изграждат в екосистемата на NVIDIA. NVIDIA Omniverse, която до 2025 г. се превърна във водещ инструмент за работа с формата USD - универсално описание на сцени, което в крайна сметка може да замени формата IFC в процесите на трансфер на информация за проектиране. Заедно с Isaac Sim, симулатор на роботизирани процеси [163], решения като NeMo Curator, Cosmos и Omniverse представляват ново ниво на автоматизация: от почистване и филтриране на данни до генериране на набор за обучение, моделиране на свойствата на обектите и обучение на роботи на строителната площадка. При това всички тези инструменти са безплатни и с отворен код, което значително намалява бариерите пред внедряването им в инженерната и строителната практика.

Автоматичната класификация на данни на ниво структурирани таблици не е толкова трудна задача, колкото изглежда на пръв поглед. Както показвахме в предишната глава (фиг. 9.1-10), при наличие на натрупани исторически данни е възможно да се компенсираат липсващите или неправилните стойности на класа въз основа на подобни параметри на други елементи. Ако елементи със сходни характеристики вече са били класифицирани правилно в няколко завършени проекта, системата може с голяма вероятност да предложи подходяща стойност за нов или незавършен елемент (фиг. 10.1-6). Подобна логика, основана на осреднени стойности и анализ на контекста, може да бъде особено ефективна при масовата обработка на таблични данни, идващи от оценки, спецификации или CAD модели.

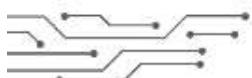


**Фиг. 10.1-6** Машинното обучение може да ви помогне автоматично да намерите средни стойности за непопълнените (бели полета) параметри на таблицата въз основа на предишни проекти.

На фона на този бърз напредък в машинното обучение е ясно, че през 2025 г. е наивно да се смята, че проблемът с автоматичното класифициране на строителни елементи ще остане дълго време нерешен. Да, настоящите алгоритми все още не са достигнали пълна зрелост, особено при непълни или хетерогенни данни, но прозорецът на възможностите за адаптиране бързо се затваря.

Компаниите, които вече инвестират в събирането, почистването и организирането на своите данни, както и в използването на инструменти за автоматизация на ETL, ще имат предимство. Останалите рискуват да изостанат - точно както някога компаниите не успяха да се справят с предизвикателствата на цифровата трансформация в транспортната и финансовата индустрия

Тези, които продължават да разчитат на ръчно управление на данните и на традиционните методи за оценка на разходите и времето, рискуват да се окажат в положението на таксиметровите автомобиларе от 2000 г., които не могат да се адаптират към ерата на мобилните приложения и автоматизираните изчисления на маршрутите до началото на 2020 г.



## ГЛАВА 10.2.

### ПРАКТИЧЕСКО РЪКОВОДСТВО ЗА ПРИЛАГАНЕ НА ПОДХОД, ОСНОВАН НА ДАННИ

#### От теория към практика: пътна карта за цифрова трансформация в строителството

Строителният бранш постепенно навлиза в нова фаза на развитие, в която познатите процеси все повече се допълват, а понякога дори заместват, от цифрови платформи и прозрачни модели на взаимодействие. Това изправя компаниите не само пред предизвикателства, но и пред значителни възможности. Онези организации, които още днес изграждат дългосрочна цифрова стратегия, ще могат не само да запазят позициите си на пазара, но и да ги разширят, като предлагат на клиентите си модерни подходи и надеждни, технологично подкрепени решения.

Важно е да се разбере, че познаването на концепциите и технологиите е само отправна точка. Мениджърите и специалистите са изправени пред практическия въпрос: откъде да започнат прилагането и как да превърнат теоретичните идеи в реална стойност. Освен това все по-често възниква въпросът: на какво ще се основава бизнесът, ако традиционните методи за изчисляване на разходите и сроковете могат да бъдат преразгледани от клиента по всяко време.

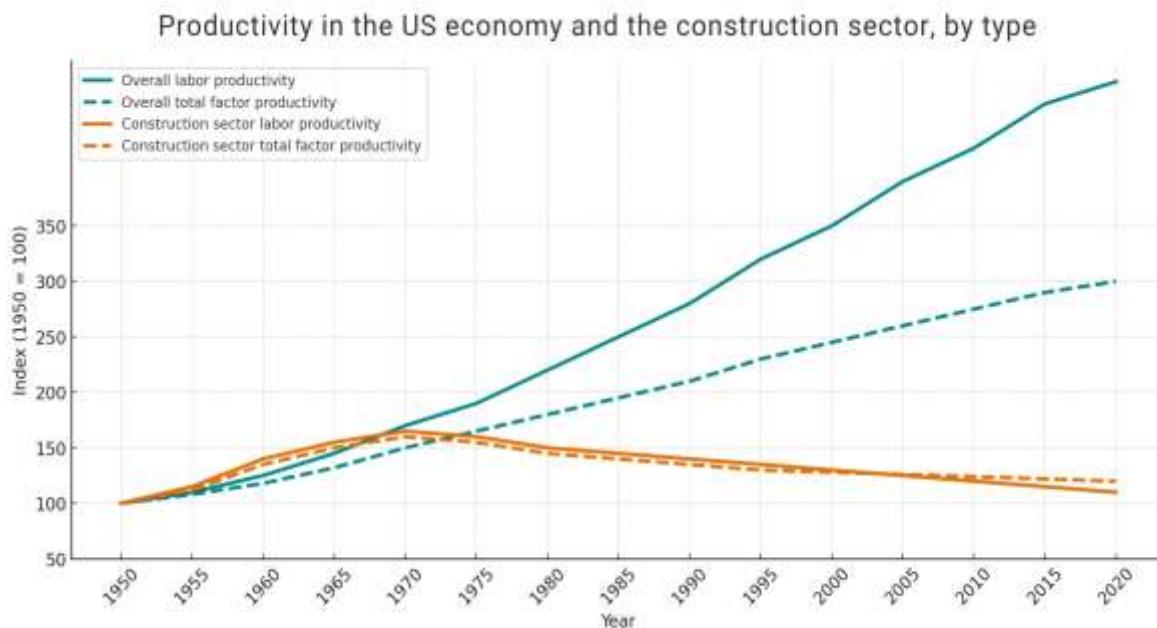
Отговорът вероятно се крие не толкова в технологиите, колкото във формирането на нова професионална култура, при която работата с данни се възприема като неразделна част от ежедневната практика. Именно липсата на внимание към цифровите технологии и иновациите е привикнала строителната индустрия към сериозното изоставане, което се наблюдава през последните десетилетия [43].

Според McKinsey разходите за научноизследователска и развойна дейност в строителството са по-малко от 1% от приходите, докато в автомобилната и космическата индустрия този показател достига 3,5-4,5%. По подобен начин разходите за ИТ в строителството остават на ниво под 1% от общите приходи [107].

В резултат на това не само нивото на автоматизация, но и производителността на труда в строителството намалява и до 2020 г. строителният работник ще произвежда по-малко, отколкото преди половин век (фиг. 10.2-1)

Подобни проблеми с производителността в строителния сектор са характерни за повечето развити и развиващи се страни (производителността в строителството е намаляла в 16 от 29-те страни от ОИСР (фигура 2.2-1)) и сочат не само липсата на технологии, но и необходимостта от системни промени в самите подходи към управлението, обучението и иновациите.

Успехът на цифровата трансформация зависи не толкова от броя и наличието на инструменти, колкото от способността на организацията да преразгледат процесите си и да развият култура, отворена за промени. Ключова е не самата технология, а хората и процесите, които осигуряват нейното ефективно използване, подпомагат непрекъснатото учене и насърчават приемането на нови идеи.



Фиг. 10.2-1 Парадоксът на производителността на труда и общата производителност на ресурсите в икономиката на САЩ и строителния сектор (1950-2020 г.) (въз основа на [43]).

В първите части на книгата моделът на бизнес средата беше сравнен с горска екосистема (Фигура 2.1-2, Фигура 1.2-4, Фигура 1.3-2). В една здрава гора периодичните пожари, при цялата им разрушителна сила, играят ключова роля за дългосрочното обновяване. Те прочистват почвата от старата растителност, връщат натрупаните хранителни вещества и създават пространство за нов живот. Някои растителни видове дори са еволюирали така, че семената им да се отварят само когато са изложени на високи температури при пожар - естествен механизъм, който осигурява идеалното време за покълване.

По същия начин в бизнеса кризите могат да действат като "контролирано изгаряне", насърчавайки появата на нови подходи и компании, които не са обвързани с останали системи. Такива периоди принуждават да се изоставят неефективни практики, като освобождават ресурси за иновации. Точно както гората след пожар започва с пионерски растения, така и бизнесът след криза формира нови, гъвкави процеси, които стават основа за зряла информационна среда.

Компаниите, които успеят правилно да интерпретират тези "сигнални огньове" и да трансформират разрушителната им енергия в конструктивна промяна, ще достигнат ново ниво на ефективност - с

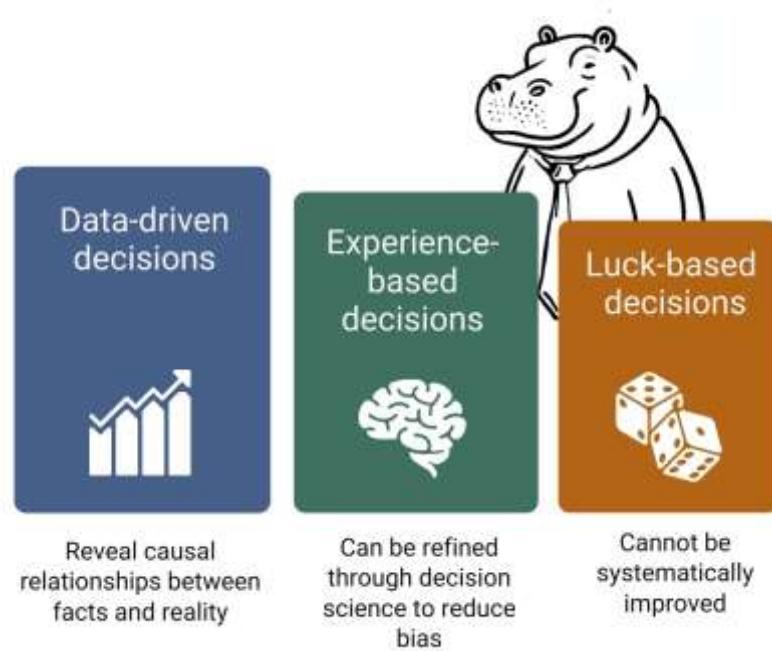
по-прозрачни и адаптивни процеси за данни, които подобряват естествената способност на организацията да се обновява и развива.

Нарастващото влияние на изкуствения интелект и машинното обучение върху бизнес средата вече не буди съмнение. Това не е просто временна тенденция, а стратегическа необходимост. Компаниите, които пренебрегват изкуствения интелект, рискуват да загубят конкурентоспособност на пазар, който все повече насищава иновациите и гъвкавостта.

Бъдещето принадлежи на тези, които възприемат ИИ не само като инструмент, но и като възможност за преосмисляне на всеки аспект от бизнеса - от оптимизирането на процесите до вземането на управлениски решения.

## Полагане на цифрова основа: 1-5 стъпки към цифрова зрялост

В тази глава разглеждаме пътната карта за цифрова трансформация и определяме основните стъпки, необходими за прилагането на подход, основан на данни, който може да помогне за трансформирането както на корпоративната култура, така и на информационната екосистема на компанията.



Фигура 10.2-2 Контролирано актуализиране и избор на стратегия: случай, опит или данни.

Според проучването на McKinsey "Зашо цифровите стратегии се провалят" (2018 г.) има поне пет причини [164], поради които компаниите не успяват да постигнат целите на цифровата трансформация

- **Несъсредоточеност:** Ръководителите и мениджърите имат различни разбириания за това

какво означава "цифрово", което води до недоразумения и несъответствия.

- **Неправилно разбиране на цифровата икономика:** много компании подценяват мащаба на промените, които цифровизацията внася в бизнес моделите и динамиката на индустрията (Фигура 10.1-6).
- **Пренебрегване на екосистемите:** компаниите се фокусират върху отделни технологични решения (силози за данни), като пренебрегват необходимостта от интегриране в по-широки цифрови екосистеми (Фигура 2.2-2, Фигура 4.1-12).
- **Подценяване на цифровизацията от страна на конкурентите:** ръководителите не вземат предвид, че конкурентите също активно внедряват цифрови технологии, което може да доведе до загуба на конкурентно предимство.
- **Пропускане на двойствения характер на цифровизацията:** изпълнителните директори делегират отговорността за цифровата трансформация на други ръководители, което бюрократизира контрола и забавя процеса на промяна.

Решаването на тези предизвикателства изисква ясно разбиране и съгласуване на цифровите стратегии на всички нива в организацията. Преди да се изгради цифрова стратегия, е важно да се разбере отправната точка. Много организации са склонни да приемат нови инструменти и платформи, без да имат пълна представа за текущото състояние.

### **Стъпка 1: Извършете одит на текущите си системи и данни.**

Преди да промените процесите, е важно да разберете какво вече е въведено. Провеждането на одит ви позволява да идентифицирате слабостите в управлението на данните и да разберете какви ресурси могат да бъдат използвани. Одитът е своеобразна "рентгенова снимка" на вашите бизнес процеси. Той ви позволява да идентифицирате рисковите области и да определите кои данни са критични за вашия проект или бизнес и кои са второстепенни.

Основни действия:

- Направете карта на ИТ средата си (в Draw.io, Lucidchart, Miro, Visio или Canva). Избройте използваните системи (ERP, CAD, CAFM, CPM, SCM и други), които участват във вашите процеси и които разглеждахме в главата "Технологии и системи за управление в съвременното строителство" (фиг. 1.2-4)
- Оценяване на въпросите, свързани с качеството на данните за всяка система, за частотата на дублиранията, възможните липсващи стойности и несъответствията във формата във всяка система.
- Идентифицирайте "болезнените точки" - места, където процесите могат да се сринат или често изискват ръчна намеса - внос, износ и допълнителни процеси на валидиране.

Ако искате екипът да се довери на отчетите, трябва да се уверите, че данните са верни от самото начало.

Одитът на качеството на данните ще ви покаже кои данни:

- Нуждае се от по-нататъшно развитие (необходимо е да се създадат процеси за автоматично почистване или допълнителни трансформации)
- Те са "боклук", който само задръства системите и от който може да се отървete, като не ги използвате повече в процесите.

Възможно е да извършите такъв одит сами. Понякога обаче е полезно да се ангажира външен консултант - особено от други отрасли: свежият поглед и независимостта от строителните "особености" ще помогнат за трезва оценка на статуквото и избягване на типичните капани на пристрастietо към определени решения и технологии.

### **Стъпка 2: Определяне на основните стандарти за хармонизиране на данните.**

След одита е необходимо да се създадат общи правила за работа с данните. Както обсъдихме в главата "Стандарти: От случайни файлове към интелигентен модел на данните", това ще помогне за премахване на изолираните потоци от данни.

Без единен стандарт всеки екип ще продължи да работи по свой собствен начин и ще поддържате "зоопарк" от интеграции, в който данните се губят при всяко преобразуване.

#### **Основни действия:**

- Изберете стандартите за данни, за да обменяте информация между системите:
  - За таблични данни това могат да бъдат структурирани формати като CSV, XLSX или по-ефективни формати като Parquet
  - За обмен на свободно структурирани данни и документи: JSON или XML
- Усвояване на работата с модели на данни:
  - Започнете с параметризиране на задачите на ниво концептуален модел на данните - както е описано в главата "Моделиране на данни: концептуален, логически и физически модел" (фиг. 4.3-2)
  - Когато навлезете по-дълбоко в логиката на бизнес процеса, преминете към формализиране на изискванията с помощта на параметри в логическия и физическия модел (фигура 4.3-6).
  - Идентифициране на ключови обекти, техните атрибути и взаимоотношения в рамките на процесите и визуализиране на тези взаимоотношения - както между обектите, така и между параметрите (фигура 4.3-7)
- Използвайте регулярни изрази (RegEx) за валидиране и стандартизиране на данни (Фигура 4.4-7), както разгледахме в главата "Структурирани изисквания и RegEx регулярни изрази". RegEx не е сложна, но изключително важна тема в работата по създаване на изисквания на ниво физически модели на данни.

Без стандарти на ниво данни и визуализация на процесите е невъзможно да се осигури последователна и мащабируема цифрова среда. Не забравяйте, че "лошите данни са скъпи". А цената на

грешката се увеличава с усложняването на проекта или организацията. Унифицирането на форматите, определянето на правила за именуване, структура и валидиране е инвестиция в стабилността и мащабируемостта на бъдещите решения.

### Стъпка 3. Внедряване на DataOps и автоматизиране на процесите.

Без добре дефинирана архитектура компаниите неминуемо ще се сблъскат с разнородни данни, съдържащи се в отделни информационни системи. Данните ще бъдат неинтегрирани, дублирани на различни места и скъпи за поддържане.

Представете си, че данните са вода, а архитектурата на данните е сложната система от тръбопроводи, която транспортира водата от източника на съхранение до мястото на използване. Архитектурата на данните определя начина, по който информацията се събира, съхранява, трансформира, анализира и предоставя на крайните потребители или приложения.

DataOps (Data Operations) е методология, която обединява събирането, почистването, валидирането и използването на данни в един автоматизиран процес, както подробно разгледахме в част 8 на книгата.

Основни действия:

- Създаване и персонализиране на ETL - конвейери за автоматизиране на процесите:
  - Извличане на данни: организирайте автоматично събиране на данни от PDF документи (фиг. 4.1-2, фиг. 4.1-5, фиг. 4.1-7), електронни таблици Excel, CAD -модели (фиг. 7.2-4), ERP -системи и други източници, с които работите
  - Трансформиране: задаване на автоматични процеси за трансформиране на данни в единен структуриран формат и автоматизиране на изчисления, които ще се извършват извън затворените приложения (фиг. 7.2-8)
  - Зареждане на: опитайте да създадете автоматично качване на данни в обобщаващи таблици, документи или централизирани хранилища (фиг. 7.2-9, фиг. 7.2-13, фиг. 7.2-16)
- Автоматизирайте процесите на изчисление и QTO (Quantity Take-Off), както обсъдихме в главата "QTO Quantity Take-Off: Групиране на данните за проекта по атрибути":
  - Настройте автоматичното извлечане на обеми от CAD -модели, като използвате API, приставки или инструменти за обратно инженерство (фиг. 5.2-5).
  - Създаване на правила за групиране на елементи за различни класове по атрибути под формата на таблици (фигура 5.2-12)
  - Опитайте се да автоматизирате често повтарящи се изчисления на обема и разходите извън модулните затворени системи (фиг. 5.2-15)
- Започнете да използвате Python и Pandas, за да обработвате данни, както разгледахме в главата "Python Pandas: незаменим инструмент за работа с данни":

- Прилагайте DataFrame за работа с XLSX файлове и автоматизиране на обработката на таблични данни (фиг. 3.4-6)
- Автоматизиране на обобщаването и преобразуването на информация чрез различни библиотеки на Python
- Използвайте LLM, за да опростите писането на готови блокове код и цели тръбопроводи (фигура 7.2-18).
- Опитайте се да създадете Pipeline на Python, който открива грешки или вижда аномалии и изпраща уведомление до отговорното лице (например ръководителя на проекта) (фигура 7.4-2)

Автоматизацията, базирана на принципите на DataOps, ви позволява да преминете от ръчна и фрагментарна обработка на данни към устойчиви и повтарящи се процеси. Това не само намалява тежестта върху служителите, които се занимават с едни и същи трансформации всеки ден, но и значително увеличава надеждността, мащабируемостта и прозрачността на цялата информационна система.

#### **Стъпка 4: Създаване на отворена екосистема за управление на данни.**

Въпреки развитието на затворените модулни системи и интегрирането им с нови инструменти, компаниите се сблъскват със сериозен проблем - нарастващата сложност на тези системи надхвърля тяхната полезност. Първоначалната идея за създаване на единна патентована платформа, обхващаща всички бизнес процеси, доведе до прекомерна централизация, при която всяка промяна изисква значителни ресурси и време за адаптиране.

Както обсъдихме в главата "Корпоративен мицел: как данните свързват бизнес процесите", ефективното управление на данните изисква отворена и единна екосистема, която свързва всички източници на информация.

Основни елементи на екосистемата:

- Изберете подходящо хранилище за данни:
  - За таблици и изчисления използвайте бази данни - например PostgreSQL или MySQL (фиг. 3.1-7).
  - За документи и отчети може да е подходящо съхранение в облак (Google Drive, OneDrive) или системи, които поддържат JSON формат.
  - Проверете възможностите на Data Warehouse, Data Lakes и други инструменти за централизирано съхранение и анализ на големи количества информация (фигура 8.1-8).
- Внедряване на решения за достъп до патентовани данни:
  - Ако използвате собствени системи, конфигурирайте достъпа до тях чрез API или SDK, за да получавате данни за външна обработка (фиг. 4.1-2)
  - Запознайте се с възможностите на инструментите за обратно инженерство за CAD формати (Фигура 4.1-13)
  - Създаване на ETL-протоколи, които периодично събират данни от приложения или

сървъри, преобразуват ги в отворени структурирани формати и ги записват в хранилища (фигура 7.2-3)

- Обсъждане в рамките на екипа на начините за осигуряване на достъп до данни без необходимост от патентован софтуер
- Запомнете: данните са по-важни от интерфейсите. Структурата и наличността на информацията, а не специфичните инструменти на потребителския интерфейс, са тези, които осигуряват дългосрочна стойност.
- Помислете за създаване на център за високи постижения (CoE) за данни, както обсъдихме в главата "Център за високи постижения (CoE) за моделиране на данни", или как можете да предоставите експертни знания за данните по други начини (фигура 4.3-9)

Екосистемата за управление на данни създава единно информационно пространство, в което всички участници в проекта работят с последователна, актуална и проверена информация. Тя е основа за мащабирами, гъвкави и надеждни цифрови процеси

## Отключване на потенциала на данните: 5-10 стъпки към цифровата зрялост

В допълнение към техническата интеграция, важен фактор за успешното прилагане на цифрови решения е тяхното приемане от крайните потребители. Ангажирането на клиентите или потребителите в измерването на ефективността е едновременно предизвикателство за подобряване на потребителското изживяване и за управление на промените в компанията. Ако дадено решение не се вписва в познатия работен процес или не решава реални проблеми на потребителите или клиентите, то няма да бъде използвано и никакви допълнителни мерки и стимули няма да поправят това.

Трансформацията е повтарящ се процес, основан на анализ на данните за взаимодействието на потребителите с новите процеси, с чести цикли на тестване, постоянна обратна връзка и усъвършенстване.

### Стъпка 5: Изграждане на култура на данните, обучение на персонала и събиране на обратна връзка

Дори и най-усъвършенстваната система няма да работи без ангажираност на служителите. Трябва да създадете среда, в която данните се използват ежедневно и екипът разбира тяхната стойност.

В публикувания през 2024 г. доклад на правителството на Обединеното кралство "Анализ на данни и изкуствен интелект при изпълнението на правителствени проекти" се отбележва [83], че обучението на специалисти с необходимите компетенции в областта на обработката и интерпретацията на данни е от решаващо значение за успешното прилагане на анализ на данни и изкуствен интелект.

Липсата на експертен опит в областта на анализа на данни е един от основните проблеми, които ограничават цифровата трансформация. Ръководителите са свикнали с установена рутина: трimesечни цикли, приоритетни инициативи и традиционни начини за придвижване на проектите. Промяната изисква отличителен лидер - достатъчно високопоставен, за да има влияние, но не толкова високопоставен, че да има време и мотивация да ръководи дългосрочен проект за трансформация.

### Основни действия:

- Признаване на необходимостта да се премине от субективни решения, основани на мнението на високоплатен служител (HiPPO), към култура на вземане на решения, основана на факти и данни, както е разгледано в главата "HiPPO или опасността от мнението при вземането на решения" (фигура 2.1-9).
- Организирайте системно обучение:
  - Провеждайте обучения за използване на структурирани данни и канете експерти от други индустрии, които не са предубедени към продукти и концепции, популярни в строителната индустрия днес.
  - Обсъждане на подходи и инструменти за анализ на данни с колеги и самостоятелно усвояване на практическа работа с инструменти като Python, pandas и LLM (фиг. 4.1-3, фиг. 4.1-6)
  - Създаване на библиотека от учебни материали (за предпочитане с кратки видеоклипове) по темата за структуриране на данни (фиг. 3.2-15) и създаване на модели на данни (фиг. 4.3-6, фиг. 4.3-7)
- Използване на съвременни технологии за обучение:
  - Използване на езикови модели (LLM) за подпомагане на манипулирането на код и данни, включително генериране на код, преработване и анализ, както и обработка и интерпретация на таблична информация (фигура 3.4-1)
  - Изследвайте как генеририаният от LLM код може да бъде адаптиран и интегриран в цялостно решение на Pipeline, когато работите в офлайн среда за разработка (IDE) (фиг. 4.4-14, фиг. 5.2-13)

Когато мениджърът продължава да взема решения по "старомодния начин", никакви обучения няма да убедят хората да приемат сериозно анализите.

Изграждането на култура на данните е невъзможно без непрекъснатата обратна връзка. Обратната връзка ви позволява да идентифицирате пропуски в процесите, инструментите и стратегията, които не могат да бъдат открити чрез вътрешни доклади или официални KPI метрики. Комплицираните коментари от потребителите на вашите решения няма да осигурят практическа стойност. Ценна е критичната обратна връзка, особено ако тя се основава на конкретни наблюдения и факти. Но получаването на такава информация изисква усилия: трябва да изградите процеси, в които участниците - както вътрешни, така и външни - да могат да споделят коментари (може да има смисъл да го правят анонимно) без изкривяване и без страх, че мненията им могат да повлият на собствената им работа. Важно е те да правят това без изкривяване и без страх от негативни последици за себе си.

Всяко обучение в крайна сметка е самообучение [165].

- Милтън Фридман, американски икономист и статистик

Внедряването на аналитични инструменти трябва да бъде придружено от редовна проверка на тяхната ефективност на практика (възвръщаемост на инвестициите, ключови показатели за ефективност), което може да се постигне само чрез структурирана обратна връзка от служителите, клиентите и партньорите. Това позволява на компаниите не само да избягват повтарянето на грешки, но и да се адаптират по-бързо към промените в средата. Наличието на механизъм за събиране и анализиране на обратна връзка е един от признаките за зрялост на организацията, която преминава от случайни цифрови инициативи към устойчив модел на непрекъснато усъвършенстване (фигура 2.2-5).

### Стъпка 6: От пилотни проекти към разширяване на обхвата

*Изберете достатъчно големи битки, за да имате значение, и достатъчно малки, за да ги спечелите.*

- Джонатан Козол

Стартирането на цифрова трансформация "наведнъж и навсякъде" е изключително рисковано. По-ефективният подход е да се започне с пилотни проекти и постепенно да се разширяват успешните практики.

Основни действия:

■ Изберете подходящия проект за пилотния проект:

- Определяне на конкретна бизнес цел или процес с измерими резултати (KPI, ROI) (фиг. 7.1-5)
- Изберете процес за автоматизация на ETL, като например автоматично валидиране на данни или изчисляване на работното натоварване (QTO) с помощта на Python и Pandas (фигура 5.2-10)
- Определяне на ясни показатели за успех (например намаляване на времето за изготвяне на спецификации за инспекция или доклади за валидиране на данни от една седмица на един ден).

■ Използвайте итеративни подходи:

- Започнете с прости процеси на преобразуване на данни и създайте поточни преобразования на многоформатни данни във форматите, от които се нуждаете за вашите процеси (фиг. 4.1-2, фиг. 4.1-5)
- Постепенно увеличавайте сложността на задачите и разширявайте автоматизацията на процесите, като формирате цялостен конвейер в IDE на базата на документирани блокове код (фиг. 4.1-7, фиг. 7.2-18).
- Документирайте и записвайте (за предпочтение с кратки видеоклипове) успешни

решения и ги споделяйте с колеги или в професионални общности.

- Разработване на шаблони и съпътстваща документация за възпроизвеждане на такива решения, така че да могат да се използват ефективно от вашите колеги (или членове на професионалната общност, включително потребители на социални медии).

Поетапното "разгръщане" ви позволява да запазите високото качество на промените и да не изпадате в хаоса на паралелните реализации. Стратегията "от малко към голямо" свежда до минимум рисковете и ви позволява да се учате от малки грешки, без да позволявате те да прераснат в критични проблеми.

Преминаването от проектен подход, при който служителите участват само частично, към формиране на постоянни екипи (напр. центрове за експертиза - ЦЕ) помага да се осигури устойчиво развитие на продукта дори след пускането на първата му версия. Такива екипи не само поддържат съществуващите решения, но и продължават да ги подобряват.

Това намалява зависимостта от продължителни одобрения: членовете на екипа са оправомощени да вземат решения в рамките на своята сфера на отговорност. В резултат на това мениджърите са освободени от необходимостта да микроуправляват, а екипите могат да се съредоточат върху създаването на реална стойност.

Разработването на нови решения не е спринт, а маратон. В него успяват тези, които първоначално се стремят към дългосрочна и последователна работа.

Важно е да се разбере, че технологиите изискват постоянно развитие. Инвестирането в дългосрочното развитие на технологичните решения е в основата на успешните операции.

### **Стъпка 7: Използване на отворени формати за данни и решения**

Както обсъдихме в главите за модулните платформи (ERP, PMIS, CAFM, CDE и т.н.), важно е да се съредоточим върху отворени и универсални формати за данни, които осигуряват независимост от решенията на доставчиците и увеличават достъпността на информацията за всички заинтересовани страни.

Основни действия:

- Преминаване от затворени формати към отворени формати:
  - Използвайте отворени формати вместо патентовани формати или намерете начин да настроите автоматично качване или конвертиране на затворени формати в отворени формати (Фигура 3.2-15).
  - Внедряване на инструменти за работа с Parquet, CSV, JSON, XLSX, които са стандарти за обмен между повечето съвременни системи (фиг. 8.1-2)
  - Ако работата с 3D геометрия играе важна роля във вашите процеси, помислете за използване на отворени формати като USD, glTF, DAE или OBJ (Фигура 3.1-14).
- Използване на векторни бази данни бази данни за ефективно анализиране и търсене на информация:

- Използвайте Bounding Box и други методи за опростяване на 3D -геометрията (Фигура 8.2-1).
- Помислете къде можете да приложите векторизация на данни - преобразуване на текстове, обекти или документи в цифрови представления (фигура 8.2-2).
- Прилагане на инструменти за анализ на големи масиви от данни:
  - Организиране на съхранението на натрупаните исторически данни (напр. PDF, XLSX, CAD) в подходящи формати за анализ (Apache Parquet, CSV, ORC) (фиг. 8.1-2).
  - Започнете да прилагате основни статистически методи и да работите с представителни извадки - или като минимум се запознайте с основните принципи на статистиката (фигура 9.2-5).
  - Внедряване и усвояване на инструменти за визуализиране на данни и на връзките между тях с цел визуализиране на резултатите от анализа. Без добра визуализация е невъзможно да се разберат напълно самите данни или процесите, базирани на тях (фигура 7.1-4).

Преминаването към отворени формати на данни и въвеждането на инструменти за анализ, съхранение и визуализиране на информация полагат основите на устойчиво и независимо цифрово управление. Това не само намалява зависимостта от доставчиците, но и осигурява равен достъп до данните за всички заинтересовани страни.

### Стъпка 8. Започнете да прилагате машинно обучение за прогнозиране

Много компании са натрупали огромни количества данни - един вид "информационни гейзери", които все още не се използват. Тези данни са били събрани в стотици или хиляди проекти, но често са били използвани само веднъж или изобщо не са били използвани в по-нататъшните процеси. Документите и моделите, съхранявани в затворени формати и системи, често се възприемат като остарели и безполезни. В действителност обаче те са ценен ресурс - основа за анализ на допуснатите грешки, автоматизиране на рутинни операции и разработване на иновативни решения за автоматична класификация и разпознаване на признания в бъдещи проекти.

Основното предизвикателство е да се научим как да извлечем тези данни и да ги превърнем в полезни прозрения. Както е разгледано в главата за машинно обучение и прогнози, техниките за машинно обучение имат потенциала значително да подобрят точността на оценките и прогнозите в различни процеси, свързани със строителството. Пълноценното използване на натрупаните данни отваря пътя към подобряване на ефективността, намаляване на риска и изграждане на устойчиви цифрови процеси.

Основни действия:

- Започнете с прости алгоритми:
  - Опитайте да приложите линейна регресия - като използвате подсказки от LLM - за прогнозиране на повтарящи се резултати в набори от данни, в които зависимостите от голям брой фактори отсъстват или са минимални (фигура 9.3-4)
  - Помислете на кои етапи от вашите процеси теоретично може да се приложи алгоритъмът k-близки съседи (k-NN) - например за задачи за класификация, оценка на

сходството на обектите или прогнозиране въз основа на исторически аналоги (фиг. 9.3-5).

■ Събиране и структуриране на данни за обучение на модели:

- Събиране на исторически данни за проекта на едно място и в един формат (фигура 9.1-10)
- Работете върху качеството и представителността на учебните образци чрез автоматизиран ETL (фигура 9.2-8).
- Научете се да разделяте данните на обучителни и тестови набори, както направихме в примера с набора от данни "Титаник" (Фигура 9.2-6, Фигура 9.2-7).

■ Разгледайте възможността за разширяване на приложението на техниките за машинно обучение към широк кръг от задачи - от прогнозиране на сроковете на проектите до оптимизиране на логистиката, управление на ресурсите и ранно идентифициране на потенциални проблеми.

Машинното обучение е инструмент за превръщане на архивираните данни в ценен актив за прогнозиране, оптимизиране и вземане на информирани решения. Започнете с малки набори от данни (Фигура 9.2-5) и прости модели, като постепенно увеличавате сложността им.

### Стъпка 9. Интегриране на IoT и съвременни технологии за събиране на данни

Светът на строителството бързо се превръща в дигитален: всяка строителна снимка, всеки пост в екипа вече е част от по-голям процес на параметризиране и маркиране на реалността. Точно както някога GPS трансформира логистиката, IoT, RFID и технологиите за автоматично събиране на данни променят строителната индустрия. Както е разгледано в главата "IoT интернет на нещата и интелигентни договори", цифровата строителна площадка със сензори и автоматизиран мониторинг е бъдещето на индустрията.

Основни действия:

- Внедряване на IoT -устройства, RFID -етикети и подробно описание на свързаните с тях процеси:
- Преценка кои области или фази на проекта могат да се възползват от най-голяма възвръщаемост на инвестициите (ROI) - например за наблюдение на температурата, вибрациите, влажността или движението на.
  - Обмислете използването на RFID за проследяване на материали, инструменти и оборудване по цялата верига на доставки.
  - Обмислете как събранныте данни могат да бъдат интегрирани в единна информационна система, като например Apache NiFi, за автоматизирана обработка и анализ в реално време (фигура 7.4-5)

■ Създаване на система за наблюдение в реално време:

- Разработване на информационни табла за проследяване на ключови показатели на процеса или проекта с помощта на инструменти за визуализация, като Streamlit, Flask или Power BI)

- Създаване на автоматични известия за сигнализиране на критични отклонения от плана или нормите (фиг. 7.4-2)
- Оценка на потенциала за прогнозна поддръжка на оборудването въз основа на събранныте данни и идентифицираните модели (фигура 9.3-6)
- Комбиниране на данни от различни източници:
  - Започнете с визуализиране на модела на данните на физическо ниво - отразете структурата на информационните потоци и ключовите параметри, идващи от CAD системи, IoT устройства и ERP платформи (фиг. 4.3-1)
  - Започнете със създаването на груб план на единна платформа за анализ на данни и подпомагане на вземането на управлениски решения. Заснемете ключовите функции, източниците на данни, потребителите и очакваните сценарии за приложение (фигура 4.3-7).

Колкото по-бързо започнете да свързвате реалните процеси с цифровия свят, толкова по-скоро ще можете да ги управлявате с данни - ефективно, прозрачно и в реално време.

#### **Стъпка 10. Подгответе се за бъдещите промени в индустрията**

Строителните компании са подложени на постоянно натиск от външната среда: икономически кризи, технологични скокове, регуляторни промени. Подобно на гора, която трябва да издържа на дъжд, сняг, суза и палещо слънце, компаниите живеят в условия на непрекъсната адаптация. И точно както дърветата стават устойчиви на студ и суза благодарение на дълбоките си коренови системи, само онези организации, които имат солидна основа от автоматизирани процеси, способност да предвиждат промените и гъвкаво да адаптират стратегиите, остават жизнеспособни и конкурентоспособни.

Както беше споменато в главата "Стратегии за оцеляване: изграждане на конкурентни предимства", строителната индустрия навлиза във фаза на радикална трансформация. Взаимодействието между възложител и изпълнител се насочва към модел на убеждаване, където прозрачността, предвидимостта и цифровите инструменти заместват традиционните подходи. В тази нова реалност печелят не най-големите, а най-гъвкавите и технологично зрелите.

Основни действия:

- Анализирайте уязвимостите на бизнеса в контекста на отворените данни:
  - Преценете как демократизирането на достъпа до данни като част от Uberisation може да има опустошително въздействие върху конкурентните ви предимства и вашия бизнес (фигура 10.1-5)
  - Помислете за стратегия за преминаване от непрозрачни и изолирани процеси към бизнес модели, основани на отворени решения, оперативна съвместимост на системите и прозрачност на данните (фиг. 2.2-5).
- Разработване на дългосрочна цифрова стратегия:
  - Определете дали се стремите да бъдете лидер в областта на иновациите или предпочитате сценарий на "догонване", при който ще пестите ресурсите си.

- Опишете етапите: краткосрочен (автоматизация на процесите, структуриране на данните), средносрочен (прилагане на LLM и ETL), дългосрочен (цифрови екосистеми, централизирани хранилища).
- Помислете за разширяване на портфолиото си от услуги:
  - Обмислете предлагането на нови услуги (насочени към енергийната ефективност, ESG, услуги за данни). Ще говорим повече за новите бизнес модели в следващата глава.
  - Стремете се да се позиционирате като надежден технологичен партньор, който поддържа целия жизнен цикъл на съоръжението - от проектирането до експлоатацията. Доверието във вас трябва да се основава на систематичен подход, прозрачни процеси и способност да предоставяте устойчиви технологични решения.

В условията на трансформация печелят не тези, които просто реагират на промените, а тези, които са проактивни. Гъвкавостта, откритостта и цифровата зрялост са основите на устойчивостта в изграждането на утрешния ден.

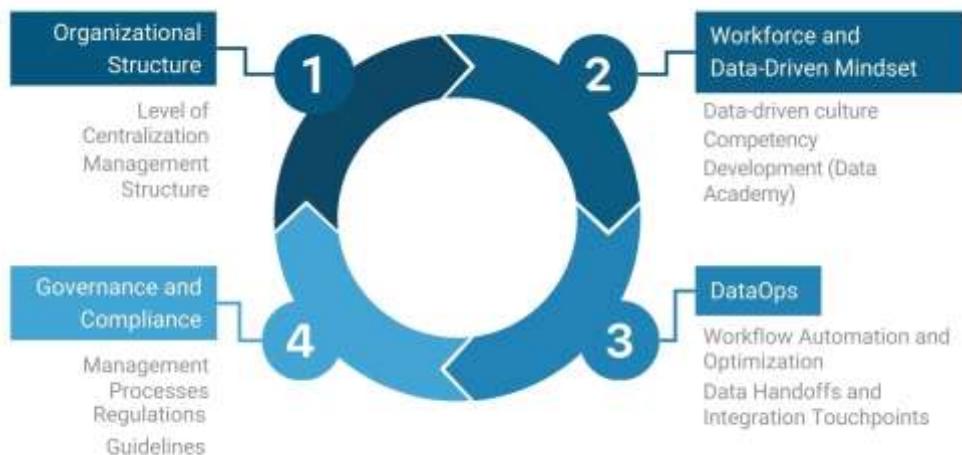
## Пътна карта за трансформация: от хаос към компания, управлявана от данни

Следващият план може да послужи като първоначален ориентир - отправна точка за оформяне на вашата собствена стратегия за цифрова трансформация, базирана на данни:

- **Одит и стандарти:** анализ на текущото състояние, унифициране на данните
- **Структуриране и класифициране на данни:** автоматизирайте преобразуването на неструктурирани и слабо структурирани данни
- **Автоматизиране на групирания,** изчисления и пресмятания: използване на инструменти и библиотеки с отворен код за автоматизация
- **Екосистема и COE:** изграждане на вътрешен екип за формиране на единна екосистема за данни в компанията
- **Култура и учене:** преминаване от HiPPO -решения към решения, базирани на данни
- **Пилотни проекти, обратна връзка и разширяване:** действайте итеративно: тествайте нови методи в ограничен мащаб, събирайте валидна обратна връзка и постепенно разширявайте решенията.
- **Отворени формати:** използвайте универсални и отворени формати, за да сте независими от доставчиците на софтуер.
- **Машинно обучение:** вграждане на ML алгоритми в процеси за прогнозиране и оптимизиране
- **IoT и цифровата строителна площадка:** интегриране на съвременните технологии за събиране на данни в процесите
- **Стратегическо адаптиране:** подготовкa за бъдещи промени в индустрията

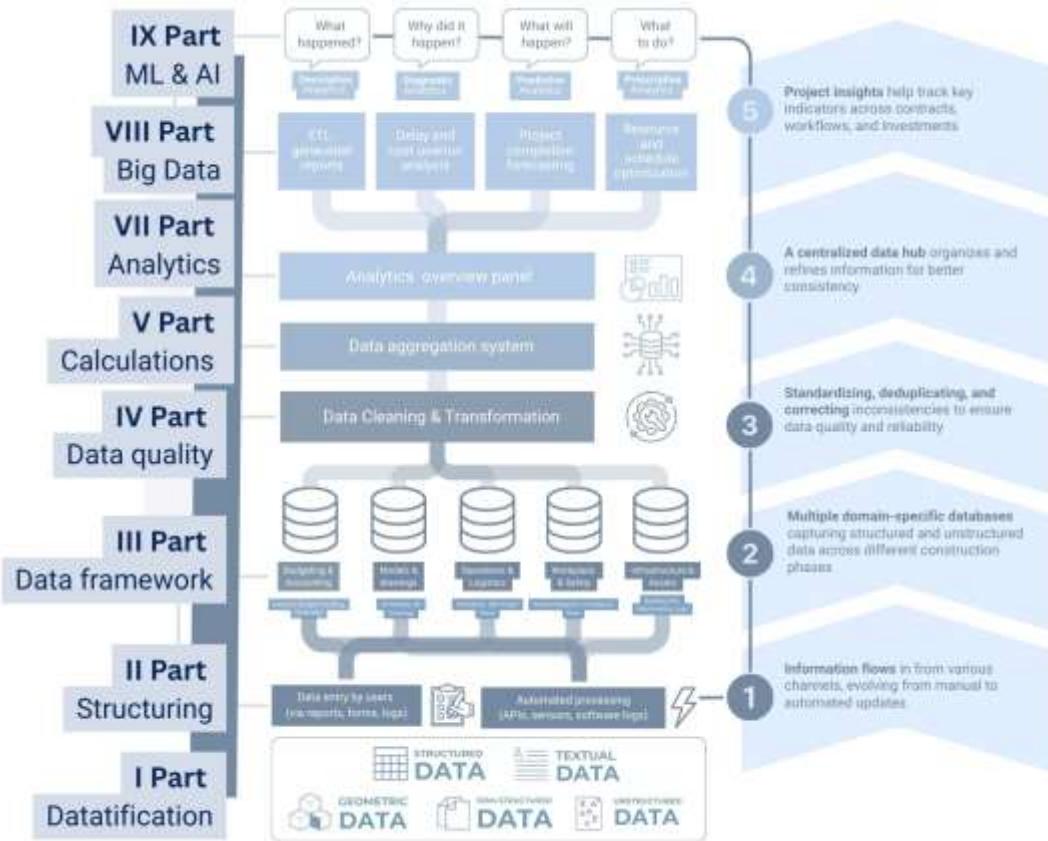
Най-важното е да не забравяте, че "данныите сами по себе си не променят компанията: променят я хората, които знаят как да работят с тези данни". Набледнете на културата, прозрачните процеси и ангажимента за непрекъснато усъвършенстване

Системният подход позволява преминаване от отделни цифрови инициативи към пълноценен модел на управление, основан на данни, при който решенията не се основават на интуиция или предположения, а на данни, факти и математически изчислени вероятности. Цифровата трансформация на строителния бранш не е свързана само с възприемането на технологии, а с оформлянето на бизнес екосистема, в която информацията за проекта се прехвърля безпроблемно и итеративно между различните системи. По този начин алгоритмите за машинно обучение осигуряват автоматичен, непрекъснат анализ, прогнозиране и оптимизиране на процесите. В такава среда спекулативните и скритите данни стават без значение - остават само доказани модели, прозрачни изчисления и предвидими резултати.



Фиг. 10.2-3 Ключови елементи на успешното управление на данни на ниво компания.

Всяка част от книгата съответства на определен етап от обработката и анализа на данни в строителните проекти (фиг. 2.2-5). Ако искате да се върнете към някоя от разгледаните по-рано теми и да я разгледате от гледна точка на цялостното разбиране на потока на използване на данните, можете да направите справка със заглавията на частите на фиг. 10.2-4.



Фиг. 10.2-4 Частите от книгата в контекста на конвейера за обработка на данни (фиг. 2.2-5): от цифровизацията на информацията до анализа и изкуствения интелект.

Независимо от размера на вашата организация, нивото на технологична зрълост или бюджета, можете да започнете да преминавате към подход, основан на данни, още днес. Дори малките стъпки в правилната посока ще дадат резултати с течение на времето.

Трансформацията, основана на данни, не е еднократен проект, а непрекъснат, повтарящ се процес на подобрене, който включва въвеждане на нови инструменти, препроектиране на процеси и развиване на култура на вземане на решения, основани на данни.

## Строителство в Индустрия 5.0: Как да печелим, когато вече не можем да се крием

Дълго време строителните компании печелеха пари от непрозрачността на процесите. Основният бизнес модел беше спекулацията - надценяване на стойността на материалите, обема на работата и процентната надценка в затворени ERP - и PMIS - системи, които са недостъпни за външни одити. Ограниченият достъп на клиентите и техните доверени лица до оригиналните данни за проекта създаде почва за схеми, при които стана почти невъзможно да се провери надеждността на изчисленията.

Този модел обаче бързо губи своята актуалност. С демократизирането на достъпа до данни, появата на LLM, навлизането на отворените данни и инструментите за автоматизация на ETL индустрията преминава към нов стандарт на работа.

В резултат на това непрозрачността вече не е конкурентно предимство - скоро тя ще се превърне в бреме, с което трудно ще се разделим. Прозрачността се превръща от опция в задължително условие за оставане на пазара.

С кого ще работят клиентите - банките, инвеститорите, физическите клиенти, клиентите на частни капиталови, правителствените клиенти - в новата цифрова реалност? Отговорът е очевиден: с тези, които са в състояние да осигурят не само резултата, но и обосновката на всяка стъпка по пътя към него. С нарастването на обема на отворените данни партньорите и клиентите ще избират компании, които гарантират прозрачност, точност и предвидимост на резултатите.

На този фон се появяват нови бизнес модели, които се основават на управление на данни и доверие, а не на спекулации:

■ **Продажба на процеси, а не на квадратни метри:** ключовият актив се превръща в доверие и ефективност, а не в конкретни споразумения с отстъпка. Основната стойност ще бъде предвидимостта на резултата въз основа на надеждни и проверени данни. Съвременните компании няма да продават строителния обект като такъв, а:

- точни крайни срокове и прозрачни работни графици;
- разумни оценки, подкрепени с изчисления;
- пълна цифрова проследимост и контрол на всички етапи на проекта.

■ **Инженеринг и анализ като услуга:** моделът "Данни като услуга" (начин за предоставяне на готови данни на потребителите чрез интернет, като услуга), при който всеки проект става част от цифрова верига от данни, а бизнес стойността е в способността за управление на тази верига. Компаниите се превръщат в интелигентни платформи, предлагящи решения, основани на автоматизация и анализ:

- Автоматизирано и прозрачно изготвяне на оценки и планове;
- оценка на риска и сроковете въз основа на алгоритми за машинно обучение;
- изчисляване на екологични показатели (ESG, CO<sub>2</sub>, енергийна ефективност);
- генериране на доклади от одитирани отворени източници.

■ **Продуциране на инженерния опит:** разработките на компанията могат да се използват многократно в рамките на компанията и да се разпространяват като отделен продукт, което представлява допълнителен източник на приходи чрез цифрови услуги. В новата среда компаниите създават не само проекти, но и цифрови активи:

- библиотеки с компоненти и шаблони за оценка;
- автоматизирани модули за проверка;
- Приставки и скриптове с отворен код (продажба на консултации) за работа с данни.

■ **Нов тип компания: Интегратор, управляван от данни:** пазарен участник, който не зависи от конкретни доставчици на софтуер или модулни системи и не е "заключен" в един-единствен софтуерен интерфейс. Той оперира свободно с данни - и изгражда своята конкурентоспособност върху тях. Строителната компания на бъдещето не е просто изпълнител, а информационен интегратор, който може да изпълнява следните функции за клиента:

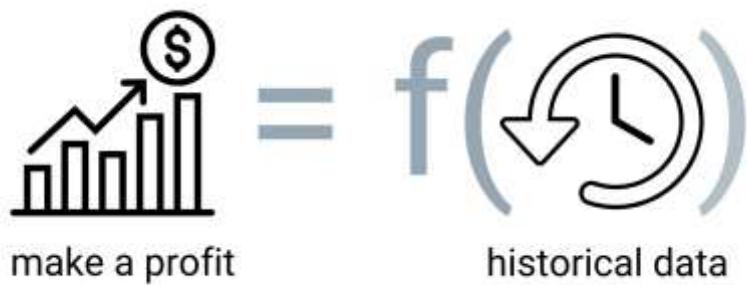
- Комбинирайте данни от различни източници и извършвайте анализи;
- Осигуряване на прозрачност и надеждност на процесите;
- консултиране за оптимизиране на бизнес процесите;
- разработване на инструменти, които работят в екосистемата за отворени данни, LLM, ETL и Pipelines.

Индустрия 5.0 (фиг. 2.1-12) бележи края на "ерата на ръчното усредняване на коефициентите" и вечерните срещи между изпълнителните директори и отдела за оценка и счетоводство. Всичко, което преди е било скрито - изчисления, оценки, обеми - става открыто, проверимо и разбирамо дори за неексперти. Тези, които първи се преориентират, ще бъдат победителите. Всички останали ще останат извън новата цифрова икономика на строителния сектор.

## ЗАКЛЮЧЕНИЕ

Строителната индустрия навлиза в епоха на фундаментални промени. Историята на управлението на информацията в строителството винаги е отразявала зрелостта на технологите на своето време - от първите записи върху глинени плочки до огромните количества цифрови данни, които се пренасят от сървърите на проектите и строителните обекти. Днес, с навлизането на автоматизацията, отворените формати и интелигентните системи за анализ, отрасът е изправен не пред постепенна еволюция, а пред бърза цифрова трансформация.

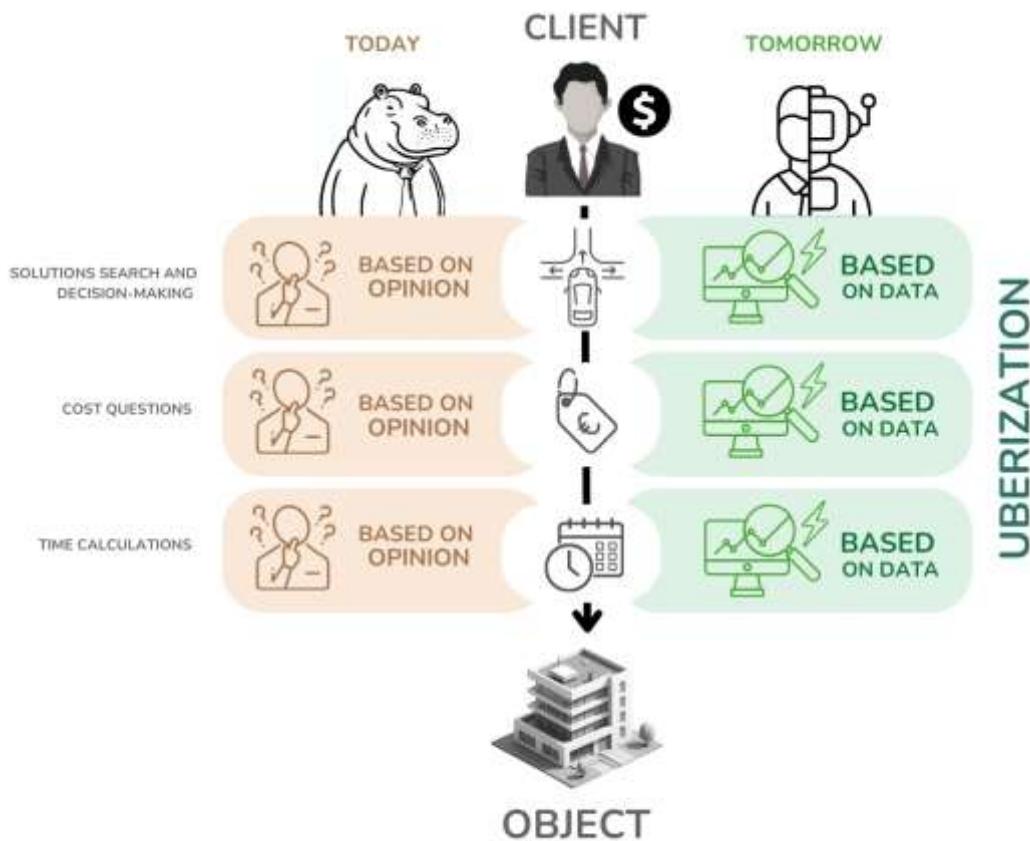
Както и в други сектори на икономиката, в строителството ще трябва да се преосмислят не само инструментите, но и принципите на работа. Фирмите, които досега диктуваха пазара и служеха като основен посредник между клиента и проекта, губят своята уникална позиция. На преден план излизат доверието и умението да се работи с данни - от събирането и структурирането им до анализирането, прогнозирането и автоматизирането на решенията.



**Фиг. 10.2-1 Структурираните исторически данни са горивото за ефективен и управляем бизнес.**

В тази книга подробно са описани основните принципи на управление на данни в строителната индустрия - от одитирането и стандартизацията до автоматизацията на процесите, използването на инструменти за визуализация и прилагането на интелигентни алгоритми. Разглеждахме как, дори с ограничени ресурси, можете да изградите работеща архитектура на данните и да започнете да вземате решения въз основа на проверими факти, а не на интуиция. Работата с данни вече не е само задача на ИТ отдела - тя се превръща в основа на управленската култура, която определя гъвкавостта, адаптивността и дългосрочната устойчивост на компанията.

Прилагането на технологиите за машинно обучение, системите за автоматична обработка, цифровите близнаци и отворените формати още днес дава възможност да се елиминира човешкият фактор там, където преди той е бил от решаващо значение. Строителството върви към автономност и управляемост, при което движението от идеята до реализацията на проекта може да се сравни с навигация в режим на автопилот: без зависимост от субективни решения, без необходимост от ръчна намеса на всеки етап, но с пълна цифрова проследимост и контрол (фиг. 10.2-2).



Фиг. 10.2-2 Преминаването от вземане на решения въз основа на мненията на важни експерти (HiPPO) към анализ на данни ще бъде настърчавано предимно от клиента.

Като усвоите методите, принципите и инструментите, представени в тази книга, ще можете да започнете да вземате в компанията си решения, основани на данни, а не на интуиция. Също така ще можете да изпълнявате вериги от модули в LLM, да копирате готови ETL конвейери във вашата среда за разработка (IDE) и да обработвате автоматично данни, за да получите необходимата ви информация в желаната от вас форма. Надграждайки върху главите от книгата, посветени на големите данни и машинното обучение, след това ще можете да реализирате по-сложни сценарии - извличане на нова информация от исторически данни и прилагане на алгоритми за машинно обучение за прогнозиране и оптимизиране на вашите процеси.

Отворените данни и процеси ще осигурят основа за по-точни оценки на разходите и сроковете на проектите, като попречат на строителните компании да спекулират с непрозрачни данни. Това е едновременно предизвикателство и възможност за индустрията да преосмисли ролята си и да се адаптира към нова среда, в която прозрачността и ефективността ще се превърнат в ключови фактори за успех.

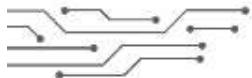
Желанието да се възприемат знанията и да се прилагат на практика е ключът към успеха в ерата на цифровата трансформация.

Компаниите, които първи осъзнават това, ще имат предимство в новата цифрова конкуренция. Но е важно да се осъзнае, че данните сами по себе си няма да променят нищо. Много хора ще трябва да

променят начина си на мислене, а това изисква стимул. Вашата компания трябва да преосмисли начина, по който споделя данни.

Хората, които променят компанията, са тези, които знаят как да работят с тези данни, да ги интерпретират, да ги използват за оптимизация и да създадат нова архитектура на процесите въз основа на тях.

Ако четете тези редове, значи сте готови за промяна и вече сте една крачка напред. Благодарим ви, че сте избрали този път. Добре дошли в ерата на цифровата трансформация!



## ЗА АВТОРА

Казвам се Артьом Бойко. Пътят ми на строителната площадка започна през 2007 г. - с работа като миньор в мина за добив на нефтени шисти в родния ми град, докато учен в Минния университет в Санкт Петербург, специалност "Минно и подземно строителство". На задната корица на тази книга можете да видите взривник в лицевата част, където добивахме и взривявахме стотици кубици нефтени шисти. Кариерата ми ме поведе в много различни посоки - от работа като работник в мините и подземното строителство до промишлен алпинист, покривник и монтажник на асансьорно оборудване. И мал съм честта да участвам в проекти от различен мащаб - от строителството на частни къщи до големи промишлени обекти в различни региони на света.



С течение на времето работата ми се измести от физическото строителство към управлението на информацията и цифровите процеси. От 2013 г. насам работя на различни позиции в малки, средни и големи строителни компании в няколко региона на Германия - от проектант до мениджър по управление на данни. Що се отнася до управлението на данни, опитът ми се състои в работа с данни в различни ERP системи, CAD (BIM), MEP, FEM, CMS. Участвал съм в оптимизация, автоматизация на процеси, както и в анализи, машинно обучение, обработка на данни във фазите на планиране, изчисляване и изпълнение на строителни работи в промишлени, жилищни, инфраструктурни и комунални строителни компании.

Работя със софтуер с отворен код и отворени данни от 2003 г. насам. През това време съм реализирал много уеб проекти - от уеб сайтове и онлайн магазини до пълноценни уеб приложения, използвайки решения с отворен код и отворени CMS. Тези платформи, които в много отношения приличат на съвременните строителни ERP системи, имат модулна архитектура, висока адаптивност и достъпност. Този опит положи основите на професионалния ми подход - фокус върху технологите с отворен код и култура на съвместна разработка. Уважението към отворения код и свободния обмен на знания е нещо, което се стремя да наследявам в строителната индустрия. Работата ми за подобряване на достъпността на данните в строителния бранш се изразява в създаването на няколко общности в социалните медии за обсъждане на отвореността на данните и използването на отворен код в строителството, както и в стартирането на няколко стартиращи компании, разработващи решения за осигуряване на достъп до данни от различни затворени системи и платформи.

Приносът ми към професионалната общност се изразява в участие като лектор в конференции, посветени на оперативната съвместимост на CAD (BIM), ERP, 4D-5D, LLM Machine Learning и Artificial Intelligence, както и в статии, публикувани в европейски издания за строителната индустрия. Едно от забележителните ми постижения е създаването на "История на BIM" [111], цялостна карта на важни софтуерни решения за управление на данни в строителната индустрия. Моята поредица от 7 статии "Разработване на BIM и лобистки игри", преведена на няколко езика, е широко призната като опит да се осветли скритата динамика на цифровите стандарти.

Така преминах от добив на скална маса към извличане и систематизиране на строителни данни. Винаги съм отворен за професионален диалог, нови идеи и съвместни проекти. Ще приема с благодарност всяка обратна връзка и ще очаквам вашите съобщения или ще ви видя сред последователите си в социалните мрежи. Много ви благодаря, че прочетохте тази книга докрай! Ще се радвам, ако тази книга ви помогне да разберете по-добре темата за данните в строителството.

## **ПО-НАТАТЪШНА ВРЪЗКА**

Мненията на читателите играят важна роля за по-нататъшното развитие на публикациите и за избора на приоритетни теми. Особено ценни са коментарите за това кои идеи са се оказали полезни и кои са породили съмнения и изискват допълнителни разяснения или цитиране на източници. Книгата включва широк спектър от материали и анализи, някои от които могат да изглеждат противоречиви или субективни. Ако в хода на четенето открите неточности, неправилно цитирани източници, логически несъответствия или печатни грешки, ще бъда благодарен за вашите коментари, мисли или критики, които можете да изпратите на адрес: [boikoartem@gmail.com](mailto:boikoartem@gmail.com). Или чрез съобщения в LinkedIn: [linkedin.com/in/boikoartem](https://linkedin.com/in/boikoartem)

Ще бъда благодарен за всяко споменаване на книгата "Data-Driven Construction" в социалните медии - споделянето на опита от четенето помага за разпространението на информация за отворените данни и инструменти и подкрепя работата ми.

## **КОМЕНТАР НА ПРЕВОДА**

Тази книга е преведена с помощта на технология за изкуствен интелект. Това направи процеса на превод много по-бърз. Въпреки това, както при всяка технологична операция, е възможно да възникнат грешки или неточности. Ако забележите нещо, което изглежда неправилно или некоректно преведено, моля, изпратете ми имейл. Вашите коментари ще помогнат за подобряване на качеството на превода.

## **СТРОИТЕЛНИ ОБЩНОСТИ, ОСНОВАНИ НА ДАННИ**

Това е място, където можете свободно да задавате въпроси и да споделяте своите проблеми и решения:

DataDrivenConstruction.io: <https://datadrivenconstruction.io>

LinkedIn: <https://www.linkedin.com/company/datadrivenconstruction/>

Twitter: <https://twitter.com/datadrivenconst>

Telegram: <https://t.me/datadrivenconstruction>

YouTube: <https://www.youtube.com/@datadrivenconstruction>

## ДРУГИ УМЕНИЯ И КОНЦЕПЦИИ

В допълнение към ключовите принципи за работа с данни в строителния бранш DataDrivenConstruction разглежда широк спектър от допълнителни концепции, програми и умения, които са от съществено значение за професионалистите, работещи с данни. Някои от тях са представени само в общ преглед, но са от решаващо значение за практиката.

Заинтересованите читатели могат да посетят уебсайта DataDrivenConstruction.io, където ще намерят връзки към допълнителни материали за ключови умения. Тези материали включват работа с Python и Pandas, изграждане на ETL -процеси, примери за обработка на данни в строителни CAD проекти, системи за големи данни и съвременни подходи за визуализация и анализ на строителни данни.

При подготовката на книгата "DataDrivenConstruction" и всички казуси са използвани много инструменти с отворен код и софтуер. Авторът би искал да благодари на разработчиците и съавторите на следните решения:

- Python и Pandas - гръбнакът на манипулирането на данни и автоматизацията
- Scipy, NumPy, Matplotlib и Scikit-Learn - библиотеки за анализ на данни и машинно обучение
- SQL и Apache Parquet - инструменти за съхраняване и обработка на големи количества строителни данни
- Отворен код CAD (BIM) инструменти за отворени данни в отворени формати
- N8n, Apache Airflow, Apache NiFi - системи за оркестрация и автоматизация на работния поток
- DeepSeek, LLaMa, Mistral - Отворен код LLM

Специални благодарности на всички участници в дискусиите по темата за отворените данни и инструментите в професионалните общности и социалните мрежи, чиито критики, коментари и идеи помогнаха за подобряване на съдържанието и структурата на тази книга.

Следете развитието на проекта на уебсайта DataDrivenConstruction.io, където се публикуват не само актуализации и корекции на книгата, но и нови глави, уроци и практически примери за прилагане на описаните техники.

## МАКСИМАЛНО УДОБСТВО С ПЕЧАТНАТА ВЕРСИЯ

Вие притежавате безплатна цифрова версия на **Data-Driven Construction**. За по-удобна работа и бърз достъп до материалите ви препоръчваме да обърнете внимание на **печатното издание**:



■ **Винаги под ръка:** книгата в печатен формат ще се превърне в надежден работен инструмент, който ще ви позволи бързо да намерите и използвате правилните визуализации и диаграми във всяка работна ситуация.

■ **Висококачествени илюстрации:** всички изображения и графики в печатното издание са представени с максимално качество.

■ **Бърз достъп до информация:** лесна навигация, възможност за правене на бележки, отметки и работа с книгата навсякъде.

Закупувайки пълната печатна версия на книгата, вие получавате удобен инструмент за комфортна и ефективна работа с информация: възможност за бързо използване на визуални материали в ежедневните задачи, бързо намиране на необходимите схеми и водене на бележки. Освен това с покупката си подкрепяте разпространението на отвореното знание.

Поръчайте печатна версия на книгата на адрес: [datadrivenconstruction.io/books](http://datadrivenconstruction.io/books)



## УНИКАЛНА ВЪЗМОЖНОСТ ЗА СТРАТЕГИЧЕСКО ПОЗИЦИОНИРАНЕ

Предлагаме ви да поставите реклами материали в безплатната версия на DataDrivenConstruction. Платената версия на изданието привлече вниманието на специалисти от повече от 50 страни по света - от Латинска Америка до Азиатско-тихоокеанския регион - още през първата година след публикуването му. За да обсъдите индивидуални условия за сътрудничество и да получите подробна информация за възможностите за пласиране, моля, попълнете формата за обратна връзка на официалния портал [datadrivenconstruction.io](http://datadrivenconstruction.io) или пишете на посочените в края на книгата контакти.



## ГЛАВИТЕ ОТ КНИГАТА СА ДОСТЪПНИ НА DATADRIVENCONSTRUCTION.IO

Можете да прочетете глави от Data-Driven Construction на уебсайта Data-Driven Construction, където постепенно се публикуват части от книгата, за да можете бързо да намерите необходимата ви информация и да я използвате в работата си. Ще намерите и много други публикации на подобни теми, както и примери за приложения и решения, които да ви помогнат да развиете уменията си и да приложите данните в строителството.



## ПОСЛЕДНИ ВЕРСИИ НА КНИГАТА ИЗТЕГЛЯНЕ ОТ ОФИЦИАЛНИЯ УЕБСАЙТ

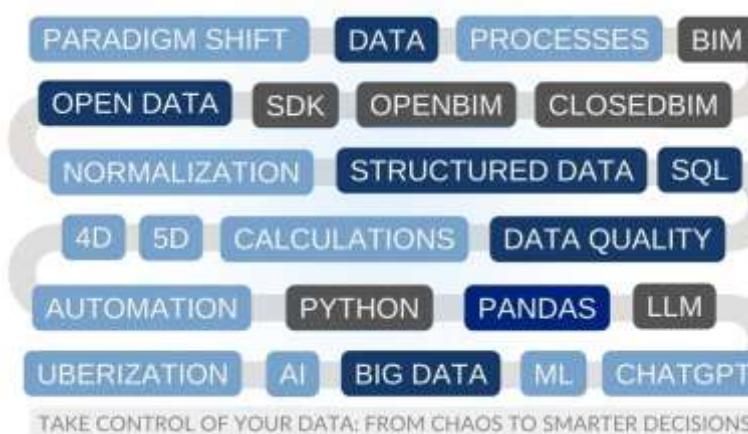
Текущите и последните версии на книгата DataDrivenConstruction са достъпни за изтегляне на адрес [datadrivenconstruction.io](http://datadrivenconstruction.io). Ако желаете да получавате актуализации с нови глави от книгата, практически съвети или прегледи на нови приложения, абонирайте се за бюллетина:

- Вие ще бъдете първите, които ще се запознаят с новите раздели на книгата
- Получаване на практически казуси и съвети за анализи и автоматизация в строителната индустрия
- Бъдете в крак с тенденциите, публикациите и примерите за приложения

Отидете на [datadrivenconstruction.io](http://datadrivenconstruction.io), за да се абонирате!

# СТРОИТЕЛСТВО, ОСНОВАНО НА ДАННИ: КОНСУЛТАЦИИ, СЕМИНАРИ И ОБУЧЕНИЕ

Програмите за обучение и консултациите на DataDrivenConstruction са помогнали на десетки водещи строителни компании по света да повишат ефективността, да намалят разходите и да подобрат качеството на решението. Сред клиентите на DataDrivenConstruction са някои от най-големите играчи на пазара с милиарди евро, включително строителни, консултантски и ИТ компании.



Защо да изберете нас?

- **Актуалност:** говорене за основните тенденции и прозрения в индустрията
- **Практически:** помага на професионалистите да решават ефикасно ежедневни задачи чрез РоС.
- **Персонализиран подход:** съобразен с вашия бизнес, като се увеличават максимално ползите от обучението и съветите

Основните области на внимание на екипа на DataDrivenConstruction:

- **Управление на качеството на данните:** помага за параметризиране на задачите, събиране на изисквания, валидиране и подготовка на данните за автоматизирана обработка.
- **Извличане на данни - извлечане и структуриране на данни:** създаване на ETL процеси и извличане на данни от имейли, PDF, Excel, изображения и други източници.
- **Анализи на BIM и CAD :** събиране, структуриране и анализиране на информация от RVT файлове, IFC, DWG и други CAD (BIM) формати.
- **Анализ и преобразуване на данни:** превръщане на разнородната информация в структурирани данни, анализи, прозрения и решения.
- **Интеграция на данни и автоматизация на процеси:** от автоматизирано създаване на документи до интеграция с вътрешни системи и външни бази данни.

Свържете се с DataDrivenConstruction.io, за да научите как използването на автоматизация може да помогне на вашата компания да постигне осезаеми бизнес резултати.

## РЕЧНИК

**AI (Artificial Intelligence)** - Изкуствен интелект; способността на компютърните системи да изпълняват задачи, които обикновено изискват човешки интелект, като разпознаване на модели, учене и вземане на решения.

**Apache Airflow** е платформа с отворен код за оркестриране на работни потоци, която ви позволява да създавате, планирате и проследявате програмно работни потоци и ETL с помощта на DAG (насочени ациклични графи).

**Apache NiFi** е инструмент за автоматизиране на потоците от данни между системите, специализиран в маршрутизирането и преобразуването на данни.

**Apache Parquet** е ефективен файлов формат за съхранение на данни в колони, оптимизиран за използване в системи за анализ на големи обеми от данни. Той осигурява значителна компресия и бърза обработка.

**API (Application Programming Interface - интерфейс за програмиране на приложения)** - формализиран интерфейс, който позволява на една програма да взаимодейства с друга без достъп до изходния код, като обменя данни и функционалност чрез стандартизираны заявки и отговори.

**Атрибут** - характеристика или свойство на обект, което описва неговите характеристики (напр. площ, обем, цена, материал).

**Базите данни** са организирани структури за съхраняване, управление и достъп до информация, използвани за ефективно извлечане и обработка на данни.

**BEP (BIM Execution Plan)** - план за внедряване на информационно моделиране на сгради, който определя целите, методите и процесите за внедряване на BIM в даден проект.

**Големи данни** - масиви от информация със значителен обем, разнообразие и скорост на обновяване, които изискват специални технологии за обработка и анализ.

**BI (Business Intelligence)** - Business Intelligence; процесите, технологиите и инструментите за трансформиране на данни в значима информация за вземане на решения.

**BIM (Building Information Modeling)** - Информационно моделиране на сгради; процесът на създаване и управление на цифрови представления на физическите и функционалните характеристики на строителните проекти, включващ не само 3D модели, но и информация за изпълнението, материалите, сроковете и разходите.

**BlackBox/WhiteBox** - подходи за разбиране на системата: в първия случай вътрешната логика е скрита, виждат се само входовете и изходите; във втория случай обработката е прозрачна и достъпна за анализ.

**Bounding Box** е геометрична конструкция, която описва границите на даден обект в триизмерното пространство чрез минимални и максимални координати X, Y и Z, създавайки "кутия" около обекта.

**BREP (Boundary Representation)** е геометрично представяне на обекти, което ги определя чрез границите на повърхностите.

**CAD (Computer-Aided Design)** е система за автоматизирано проектиране, която се използва за създаване, редактиране и анализиране на точни чертежи и 3D модели в архитектурата, строителството, инженерството и други индустрии.

**CAFM (Computer-Aided Facility Management)** е софтуер за управление на собствеността и инфраструктурата, който включва планиране на пространството, управление на активите, поддръжка и мониторинг на разходите.

**CDE (Обща среда за данни)** - централизирано цифрово пространство за управление, съхранение, споделяне и съвместна работа с информация за проекта на всички етапи от жизнения цикъл на обекта.

**Центрърът за върхови постижения (ЦВП)** е специализирана структура в рамките на една организация, която отговаря за развитието на конкретна област на знанието, разработването на стандарти и най-добри практики, обучението на персонала и подпомагането на въвеждането на иновации.

**CoClass** е модерна система за класификация на строителни елементи от трето поколение.

**Концептуалният модел на данните** е представяне на високо ниво на основните същности и техните взаимоотношения без подробности за атрибутите, което се използва в началните етапи на проектиране на базата данни.

**CRM (Customer Relationship Management - управление на взаимоотношенията с клиенти)** е система за управление на взаимоотношенията с клиенти, която се използва за автоматизиране на процесите на продажби и обслужване.

**DAG (Directed Acyclic Graph - насочен ацикличен граф)** е насочен ацикличен граф, използван в системите за оркестрация на данни (Airflow, NiFi) за определяне на последователността на задачите и зависимостите.

**Dash** е рамка на Python за създаване на интерактивни уеб базирани визуализации на данни.

**Информационно табло - Информационно табло**, което визуално представя ключови показатели за ефективност и метрики в реално време.

**Подходът, ориентиран към данните**, е методология, която дава приоритет на данните пред приложението или софтуерния код, превръщайки данните в основен актив на организацията.

**Управление на данните** - набор от практики, процеси и политики, които осигуряват подходящо и ефективно използване на данните в рамките на организацията, включително контрол на достъпа, качеството и сигурността.

**Езерото с данни** е съоръжение за съхранение, предназначено за съхраняване на големи количества необработени данни в оригиналния им формат до момента на използването им.

**Data Lakehouse** е архитектурен подход, който съчетава гъвкавостта и мащабируемостта на езерата от данни (**Data Lake**) с управляемостта и производителността на складовете за данни (DWH).

**Строителството, базирано на данни**, е стратегически подход, при който всеки етап от жизнения цикъл на обекта - от проектирането до експлоатацията - се подпомага от автоматизирани, взаимосвързани системи. Този подход осигурява непрекъснато обучение, основано на факти, намалява несигурността и позволява на компаниите да постигнат устойчиво лидерство в индустрията.

**Data-Driven Integrator** - компания, специализирана в комбинирането на данни от различни източници и анализирането им за вземане на управленски решения.

**Подход, базиран на данни** - методология, при която данните се разглеждат като стратегически актив и решенията се вземат въз основа на обективен анализ на информацията, а не на субективни мнения.

**Минимализъм на данните** - подход за намаляване на данните до най-ценните и значимите, което позволява опростена обработка и анализ на информацията.

**Блато от данни** - разпръснат масив от неструктурирани данни, който се появява, когато информацията се събира и съхранява по неконтролиран начин без подходяща организация.

**DataOps** е методология, която съчетава принципите на DevOps, данни и анализи, като се фокусира върху подобряване на сътрудничеството, интеграцията и автоматизацията на потоците от данни.

**Цифровизацията на информацията** е процесът на преобразуване на всички аспекти на строителната дейност в цифрова форма, подходяща за анализ, тълкуване и автоматизация.

**DataFrame** - Двуизмерна таблична структура от данни в библиотеката Pandas, в която редовете представляват отделни записи или обекти, а колоните - техните характеристики или атрибути.

**Дескриптивен** анализ - анализиране на исторически данни, за да се разбере какво се е случило в миналото.

**Диагностичен** анализ - анализиране на данни, за да се определи защо нещо се е случило.

**Диаграмата на Гант** е инструмент за планиране на проекти, който представя задачите като хоризонтални ленти върху времева линия, което ви позволява да визуализирате последователността и продължителността на работата.

**DWH (Data Warehouse)** е централизирана система за съхранение на данни, която обединява информация от различни източници, структурира я и я предоставя за анализ и отчитане.

**ESG (Environmental, Social, Governance)** - набор от критерии за оценка на екологичните, социалните и управленските въздействия на дадена компания или проект.

**ELT (Extract, Load, Transform - извличане, зареждане и преобразуване)** е процес, при който данните първо се извличат от източници и се зареждат в хранилище, а след това се преобразуват за аналитични цели.

**ETL (Extract, Transform, Load - извличане, преобразуване, зареждане)** е процесът на извличане на данни от различни източници, преобразуването им в желания формат и зареждането им в целевото хранилище за анализ.

**ER-диаграма (Entity-Relationship)** - визуална диаграма, показваща същностите, техните атрибути и връзките между тях, използвана при моделирането на данни.

**ERP (планиране на ресурсите на предприятието)** е цялостна модулна система за планиране на ресурсите на предприятието, която се използва за управление и оптимизиране на различни аспекти на строителния процес.

**Характеристики** - В машинното обучение независими променливи или атрибути, използвани като входни данни за модела.

**Физически модел на данните** - подробно представяне на структурата на базата данни, включително таблици, колони, типове данни, ключове и индекси, оптимизирани за конкретна СУБД.

**FPDF** е библиотека на Python за създаване на PDF документи.

**Геометричното ядро** е софтуерен компонент, който предоставя основни алгоритми за създаване, редактиране и анализиране на геометрични обекти в CAD, BIM и други инженерни приложения.

**HiPPO (Highest Paid Person's Opinion - мнение на най-високоплатеното лице)** - подход за вземане на решения, основан на мнението на най-високоплатеното лице в организацията, а не на обективни данни.

**IDE (Integrated Development Environment - интегрирана среда за разработка)** - интегрирана среда за разработка, цялостен инструмент за писане, тестване и отстраняване на грешки в кода (напр. PyCharm, VS Code, Jupyter Notebook).

**IDS (Information Delivery Specification - спецификация за предоставяне на информация)** е спецификация за предоставяне на информация, която определя изискванията към данните на различни етапи от проекта.

**IFC (Industry Foundation Classes)** е формат за обмен на данни за BIM, който осигурява оперативна съвместимост между различни софтуерни решения.

**Индустрия 5.0** е концепция за индустриално развитие, която съчетава възможностите на цифровизацията, автоматизацията и изкуствения интелект с човешкия потенциал и екологичната устойчивост.

**Интегрирането на данни** е процесът на комбиниране на данни от различни източници в една единна, съгласувана система, за да се осигури единен поглед върху информацията.

**Информационните силози** са изолирани системи за съхранение на данни, които не обменят информация с други системи, което създава пречки пред ефективното използване на данните.

**Интернет на нещата (IoT)** е концепцията за свързване на физически обекти към интернет с цел събиране, обработка и предаване на данни.

**k-NN (k-Nearest Neighbors - най-близки съседи)** е алгоритъм за машинно обучение, който класифицира обекти въз основа на сходство с най-близките съседи в обучаващата извадка.

**Kaggle** е платформа за състезания по анализ на данни и машинно обучение.

**Калкулация** - изчисляване на разходите за строителни работи или процеси за определена мерна единица (напр. 1 m<sup>2</sup> стена от гипсокартон, 1 m<sup>3</sup> бетон).

**KPI (Key Performance Indicators)** са ключови показатели за ефективност, количествено измерими показатели, използвани за оценка на успеха на дадена компания или конкретен проект.

**Етикети** - В машинното обучение целевите променливи или атрибути, които моделът трява да предскаже.

**Алгоритъм за обучение** - процесът на намиране на най-добрата хипотеза в модел, съответстващ на целева функция, като се използва набор от данни за обучение.

**Линейна регресия** - статистически метод за моделиране на връзката между зависима променлива и една или повече независими променливи.

**LLM (Large Language Model)** - Large Language Model, изкуствен интелект, обучен да разбира и генерира текст от огромни количества данни, способен да анализира контекста и да пише програмен код.

**LOD (Level of Detail/Development)** - нивото на детайлност на модела, което определя степента на геометрична точност и информационно съдържание.

**Логическият модел на данни** е подробно описание на същности, атрибути, ключове и връзки, отразяващи бизнес информация и правила, междинен етап между концептуалния и физическия модел.

**Машинно обучение** - клас техники за изкуствен интелект, които позволяват на компютърните системи да се учат и да правят прогнози от данни без изрично програмиране.

**Masterformat** е първо поколение класификационна система, използвана за структуриране на строителните спецификации в раздели и дисциплини.

**MEP (Mechanical, Electrical, Plumbing)** - Сградни инженерни системи, които включват механични, електрически и водопроводни компоненти.

**Мрежата** е мрежово представяне на 3D обекти, състоящо се от върхове, ръбове и лица.

**Модел** - В машинното обучение набор от различни хипотези, една от които апроксимира целевата функция, която трява да бъде предсказана или апроксимирана.

**Моделирането на данни** е процесът на създаване на структурирано представяне на данните и техните взаимоотношения за внедряване в информационните системи, включително на концептуално, логическо и физическо ниво.

**n8n** е инструмент с отворен код за автоматизиране на работните процеси и интегриране на приложения чрез подход с нисък код.

**Нормализация** - в машинното обучение процесът на привеждане на различни цифрови данни към обща скала, за да се улесни обработката и анализът.

**Обратното инженерство** - процесът на изучаване на устройството, функционирането и технологията на производство на даден обект чрез анализиране на неговата структура, функции и действие. В контекста на данните - извлечане на информация от патентовани формати за използване в отворени системи.

**OCR (Optical Character Recognition - оптично разпознаване на символи)** е технология за **оптично разпознаване на символи**, която преобразува текстови изображения (сканирани документи, снимки) в машиночетим текстов формат.

**OmniClass** е второ поколение международен класификационен стандарт за управление на строителната информация.

**Онтология** - система от взаимовръзки между понятия, която формализира определена област на знанието.

**Отворен код** - модел за разработване и разпространение на софтуер с отворен код, който е достъпен за свободно използване, изучаване и модифициране.

**Open BIM** е концепция за отворен BIM, която включва използването на отворени стандарти и формати за обмен на данни между различни софтуерни решения.

**Отворени стандарти** - публично достъпни спецификации за изпълнение на конкретна задача, които позволяват на различни системи да си взаимодействат и да обменят данни.

**Pandas** е библиотека на Python с отворен код за обработка и анализ на данни, която предоставя структури за данни DataFrame и Series за ефективна обработка на таблична информация.

**Парадигмата на отворените данни** е подход към обработката на данни, при който информацията се предоставя свободно за използване, повторна употреба и разпространение от всеки.

**Параметричният** метод е метод за оценка на строителни проекти, който използва статистически модели за оценка на стойността въз основа на параметрите на проекта.

**PIMS (Project Information Model)** е цифрова система, предназначена за организиране, съхраняване и споделяне на цялата информация за проекта.

**Конвейер** - Последователност от процеси за обработка на данни - от извлечане и преобразуване до анализ и визуализация.

**PMIS (Project Information Management System)** е система за управление на проекти, предназначена за подробен контрол на задачите на ниво отделен строителен проект.

**Прогнозният** анализ е раздел от анализа, който използва статистически методи и машинно обучение за прогнозиране на бъдещи резултати въз основа на исторически данни.

**Предписващ** анализ - раздел от анализа, който не само предсказва бъдещи резултати, но и предлага оптимални действия за постигане на желаните резултати.

**Патентованите** формати са затворени формати за данни, контролирани от определена компания, които ограничават възможността за обмен на информация и увеличават зависимостта от конкретен софтуер.

**QTO (Quantity Take-Off)** е процесът на извлечане на количествата на елементите от проектната документация, за да се изчислят количествата на материалите, необходими за реализацията на проекта.

**Система за управление на качеството** - система за управление на качеството, която гарантира, че процесите и резултатите отговарят на установените изисквания.

**RAG (Retrieval-Augmented Generation)** е метод, който съчетава генеративните възможности на езиковите модели с извлечането на релевантна информация от корпоративни бази данни, като подобрява точността и релевантността на отговорите.

**СУБД (Система за управление на релационни бази данни)** е система за управление на релационни бази данни, която организира информацията под формата на взаимосвързани таблици.

**RegEx (Regular Expressions)** е формализиран език за търсене и обработка на низове, който позволява да се задават шаблони за проверка на текстови данни за съответствие с определени критерии.

**Регресията** е статистически метод за анализ на връзката между променливите.

**Изчисленията на CO<sub>2</sub>** са метод за оценка на емисиите на въглероден диоксид, свързани с производството и използването на строителни материали и процеси.

**Методът на ресурсите** е метод за изготвяне на оценки въз основа на подробен анализ на всички ресурси (материали, труд, оборудване), необходими за извършване на строителните работи.

**RFID (Radio Frequency Identification)** е технология за автоматично идентифициране на обекти чрез радиосигнали, използвана за проследяване на материали, машини и персонал.

**Възвръщаемост на инвестициите (ROI)** е показател, отразяващ съотношението между печалбата и инвестираните средства, който се използва за оценка на ефективността на инвестициите.

**SaaS (софтуер като услуга)** е модел на софтуер като услуга, при който приложенията се хостят от доставчик и се предоставят на потребителите по интернет.

**SCM (Supply Chain Management)** - управление на веригата за доставки, което включва координиране и оптимизиране на всички процеси от снабдяването с материали до доставката на готовите продукти.

**Силозите от данни** са изолирани хранилища на информация в рамките на една организация, които не са интегрирани с други системи, което затруднява обмена на данни и го прави неефективен.

**SQL (Structured Query Language)** е структуриран език за заявки, използван за работа с релационни бази данни.

**SQLite** е лека, вграждаща се, междуплатформена СУБД, която не изисква отделен сървър и поддържа основни функции на SQL, широко използвана в мобилни приложения и вградени системи.

**Структурирани данни** - информация, организирана в определен формат с ясна структура, например в релационни бази данни или таблици.

**Свободно структурирани данни** - информация с частична организация и гъвкава структура, като JSON или XML, където различните елементи могат да съдържат различни набори от атрибути.

Същност е конкретен или абстрактен обект от реалния свят, който може да бъде еднозначно идентифициран, описан и представен под формата на данни.

**Контролирано обучение** - вид машинно обучение, при което алгоритъмът се обучава върху разделени данни, при които желаният резултат е известен за всеки пример.

**Таксономията** е йерархична класификационна система, използвана за систематично категоризиране на елементи въз основа на общи характеристики.

**Наборът от данни Titanic** е популярен набор от данни за обучение и тестване на модели за машинно обучение.

**Обучение** - Процесът, при който алгоритъмът за машинно обучение анализира данни, за да идентифицира модели и да създаде модел.

**Трансферното обучение** е техника за машинно обучение, при която модел, обучен за една задача, се използва като отправна точка за друга задача.

**Трансформиране** - Процесът на промяна на формата, структурата или съдържанието на данните с цел последващо използване.

**Изисквания към данните** - формализирани критерии, определящи структурата, формата, пълнотата и качеството на информацията, необходима за подпомагане на бизнес процесите.

**Уберизацията на строителния бранш** е процесът на трансформация на традиционните бизнес модели в строителството под влиянието на цифрови платформи, които осигуряват пряко взаимодействие между клиентите и изпълнителите без посредници.

**Uniclass** е система за класификация на строителни елементи от второ и трето поколение, широко използвана в Обединеното кралство.

**USD (Universal Scene Description - универсално описание на сцената)** е формат за данни, разработен за компютърна графика, но намира приложение в инженерните системи поради простата си структура и независимостта си от геометрични ядра.

**Валидирането на данните** е процесът на проверка на информацията спрямо установени критерии и изисквания, за да се гарантира точността, пълнотата и последователността на данните.

**Векторна база данни** - Специализиран тип база данни, която съхранява данни като многомерни вектори за ефективно семантично търсене и сравняване на обекти.

**Векторното представяне (вграждане)** е метод за преобразуване на данни в многомерни числови вектори, който позволява на машинните алгоритми ефективно да обработват и анализират информация.

**VectorOps** е методология, насочена към обработката, съхранението и анализа на многомерни векторни данни, особено важна в области като цифровите близнacci и семантичното търсене.

**Визуализация** - графично представяне на данни за по-добро възприемане и анализ на информациите.

Азбучната категоризация на термините е направена по техните английски имена.

## СПИСЪК С РЕФЕРЕНЦИИ И ОНЛАЙН МАТЕРИАЛИ

- [1] Gartner, "IT Key Metrics Data 2017: Index of Published Documents and Metrics" (Данни за ключови показатели за ИТ 2017 г.: Индекс на публикуваните документи и показатели), 12 декември 2016 г. [Онлайн]. Достъпно на: <https://www.gartner.com/en/documents/3530919>. [Дата на достъп: 1 март 2025 г.].
- [2] KPMG, "Познати предизвикателства - нови подходи. Глобално проучване на строителството за 2023 г.", 1 януари 2023 г. [Онлайн]. Достъпно на: <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2023/06/familiar-challenges-new-solutions-1.pdf>. [Достъп: 5 март 2025 г.].
- [3] F. R. Барнард, "Една снимка струва хиляда думи", 10 Мар 1927. [Онлайн]. Достъпно на: [https://en.wikipedia.org/wiki/A\\_picture\\_is\\_worth\\_a\\_thousand\\_words](https://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words). [Дата на обръщението: 15 март 2025 г.].
- [4] М. Бестиан, "Главният изпълнителен директор на Microsoft Сатя Надела казва, че самопровъзгласяването на AGI е "безсмислено хакерство", 21 февруари 2025 г. [Онлайн]. Достъпно на: <https://the-decoder.com/microsoft-ceo-satya-nadella-says-self-claimingagi-is-nonsensical-benchmark-hacking/>. [Дата на достъп: 15 март 2025 г.].
- [5] W. E. Forum, "Форумът определя бъдещето на строителството - пейзаж в трансформация", 1 януари 2016 г. [Онлайн]. Достъпно на: [https://www3.weforum.org/docs/WEF\\_Shaping\\_the\\_Future\\_of\\_Construction.pdf](https://www3.weforum.org/docs/WEF_Shaping_the_Future_of_Construction.pdf). [Дата на достъп: 2 март 2025 г.].
- [6] С. Д. Гилеспи, "Глината: заплитането на Земята в епохата на глината", 2024 г. [Онлайн]. Достъпно на: <https://ufl.pb.unizin.org/imos/chapter/clay/>.
- [7] "Папирус от III в. пр.н.е. Езикът е гръцки", 2024 г. [Онлайн]. Available: <https://www.facebook.com/429710190886668/posts/595698270954525>.
- [8] "Мониторинг: използване на наличните инструменти", 1980 г. [Онлайн]. Достъпно на: <https://pubmed.ncbi.nlm.nih.gov/10246720/>. [Дата на обръщението: 15 март 2025 г.].
- [9] PWC, "Data driven What students need to succeed in a rapidly changing business world" ("Данни, от които учениците се нуждаят, за да успеят в бързо променящия се бизнес свят"), 15 февруари 2015 г. [Онлайн]. Достъпно на: <https://www.pwc.com/us/en/faculty-resource/assets/PwC-Data-driven-paper-Feb2015.pdf>. [Дата на достъп: 15 март 2025 г.].
- [10] Skanska USA, "Тенденции на пазара на есенното строителство", 2 ноември 2023 г. [Онлайн]. Available: <https://x.com/SkanskaUSA/status/1720167220817588714>.
- [11] "Оксфордски основни цитати" (4-то издание), Oxford University Press, 2016 г. [Онлайн].

- Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191826719.001.0001/q-oro-ed4-00006236>. [Дата на достъп: 1 март 2025 г.].
- [12] "Цитат: Сондергаард за анализа на данни," [Онлайн]. Достъпно на: <https://www.causeweb.org/cause/resources/library/r2493>. [Дата на достъп: 15 март 2025 г.].
- [13] "Как глобалният интерес към изкуствения интелект стимулира пазара за управление на данни", 28 май 2024 г. [Онлайн]. Достъпно на: <https://iot-analytics.com/how-global-ai-interest-is-boosting-data-management-market/>. [Дата на достъп: 15 март 2025 г.].
- [14] И. McCue, "История на ERP", 2024. [Онлайн]. Достъпно на: <https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>.
- [15] erpscout, "Цена на ERP: колко струва една ERP система?", [Онлайн]. Достъпно на: <https://erpscout.de/en/erp-costs/>. [Дата на достъп: 15 март 2025 г.].
- [16] softwarepath, "Какво ни казват 1384 ERP проекта за избора на ERP (доклад за ERP за 2022 г.)", 18 януари 2022 г. [Онлайн]. Достъпно на: <https://softwarepath.com/guides/erp-report>. [Дата на достъп: 15 март 2025 г.].
- [17] Deloitte, "Data-Driven Management in Digital Capital Projects", 16 декември 2016 г. [Онлайн]. Достъпно на: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/Real%20Estate/us-engineering-construction-data-driven-management-digital-capital-projects.pdf>. [Дата на обръщение: 1 март 2025 г.].
- [18] McKinsey, "Предприятието, управлявано от данни, през 2025 г.", 28 януари 2022 г. [Онлайн]. Достъпно на: <https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-data-driven-enterprise-of-2025>. [Дата на справка: 22 май 2024 г.].
- [19] Wikipedia, "Законът на Мур". [Онлайн]. Достъпно на: [https://en.wikipedia.org/wiki/Moore%27s\\_law](https://en.wikipedia.org/wiki/Moore%27s_law). [Дата на достъп: 15 март 2025 г.].
- [20] Accenture, "Изграждане на по-голяма стойност с капиталовите проекти", 1 януари 2020 г. [Онлайн]. Достъпно на: <https://www.accenture.com/content/dam/accenture/final/a-com-migration/r3-3/pdf/pdf-143/accidenture-industryx-building-value-capital-projects-highres.pdf>. [Дата на достъп: 3 март 2024 г.].
- [21] Б. Mapp, "Колко данни създаваме всеки ден? Умопомрачителните статистики, които всеки трябва да прочете", 2018 г. [Онлайн]. Достъпно на: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read>.
- [22] "Колко данни се създават всеки ден?", 2024 г. [Онлайн]. Достъпно на: <https://graduate.northeastern.edu/resources/how-much-data-produced-every-day/>.
- [23] Т. Съливан, "ИИ и глобалната "сфера на данните": колко информация ще има човечеството

- до 2025 г.?”, 2024 г. [Онлайн]. Достъпно на: <https://www.datauniverseevent.com/en-us/blog/general/AI-and-the-Global-Datasphere-How-Much-Information-Will-Humanity-Have-By-2025.html>.
- [24] Statista, "Общ брой на печатните книги, произведени в различни региони на Западна Европа през всяко полугодие между 1454 и 1800 г.", [интернет]. [Онлайн]. Достъпно на: <https://www.statista.com/statistics/1396121/europe-book-production-half-century-region-historical/>. [Дата на справка: 1 март 2025 г.].
- [25] "Примери за ценообразуване", 2024 г. [Онлайн]. Достъпно на: <https://cloud.google.com/storage/pricing-examples>.
- [26] М. Ашаре, "Предприятията възлагат съхранението на данни на външни изпълнители, тъй като сложността нараства", 10 май 2024 г. [Онлайн]. Достъпно на: <https://www.ciodive.com/news/enterprises-outsource-data-storage-complexity-rises/715854/>. [Дата на достъп: 15 март 2025 г.].
- [27] JETSOFTPRO, "SaaS е мъртъв? Шокиращата прогноза на главния изпълнителен директор на Microsoft е обяснена", 13 януари 2025 г. [Онлайн]. Достъпно на: <https://jetsoftpro.com/blog/saas-is-dead/>.
- [28] BG2 Pod, "Сатя Надела | BG2 с Бил Гърли и Брад Герстнер", 12 декември 2024 г. [Онлайн]. Достъпно на: [https://www.youtube.com/watch?v=9NtsnzRFJ\\_o](https://www.youtube.com/watch?v=9NtsnzRFJ_o). [Дата на достъп: 15 март 2025 г.].
- [29] GoodReads, "Тим Бърнърс-Лий", [Онлайн]. Достъпно на: <https://www.goodreads.com/quotes/8644920-data-is-a-precious-thing-and-will-last-longer-than>. [Дата на достъп: 15 март 2025 г.].
- [30] KPMG, "Cue Construction 4.0: Време за промяна", 1 януари 2023 г. [Онлайн]. Достъпно на: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Достъп: 5 март 2025 г.].
- [31] И. Дейнингер, Б. Кох, Р. Баукнехт и М. Лангханс, "Използване на цифрови модели за декарбонизация на производствен обект: Пример за свързване на модел на сграда, производствен модел и енергиен модел", 2024. [Онлайн]. Достъпно на: [https://www.researchgate.net/publication/374023998\\_Using\\_Digital\\_Models\\_to\\_Decarbonize\\_a\\_Production\\_Site\\_A\\_Case\\_Study\\_of\\_Connecting\\_the\\_Building\\_Model\\_Production\\_Model\\_and\\_Energy\\_Model](https://www.researchgate.net/publication/374023998_Using_Digital_Models_to_Decarbonize_a_Production_Site_A_Case_Study_of_Connecting_the_Building_Model_Production_Model_and_Energy_Model).
- [32] McKinsey, "REINVENTING CONSTRUCTION: A ROUTE TO HIGHER PRODUCTIVITY", 1 февруари 2017 г. [Онлайн]. Достъпно на: <https://www.mckinsey.com/~/media/mckinsey/business%20functions/operations/our%20insights/reinventing%20construction%20through%20a%20productivity%20revolution/mgi-reinventing-construction-a-route-to-higher-productivity-full-report.pdf>.

- [33] Работна група по въпросите на строителството до заместник министър-председателя, "Преосмисляне на строителството", 1 октомври 2014 г. [Онлайн]. Достъпно на: [https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking\\_construction\\_report.pdf](https://constructingexcellence.org.uk/wp-content/uploads/2014/10/rethinking_construction_report.pdf).
- [34] Forbes, "Без мнение сте просто още един човек с данни", 15 март 2016 г. [Онлайн]. Достъпно на: <https://www.forbes.com/sites/silberzahnjones/2016/03/15/without-an-opinion-youre-just-another-person-with-data/>. [Дата на достъп: 15 март 2025 г.].
- [35] Уикицитат, "Чарлз Бабидж", [Онлайн]. Достъпно на: [https://en.wikiquote.org/wiki/Charles\\_Babbage](https://en.wikiquote.org/wiki/Charles_Babbage). [Дата на достъп: 15 март 2025 г.].
- [36] SAP, "Ново проучване установява, че почти половината от ръководителите се доверяват на изкуствения интелект повече от самите себе си", 12 март 2025 г. [Онлайн]. Достъпно на: <https://news.sap.com/2025/03/new-research-executive-trust-ai/>. [Дата на достъп: 15 март 2025 г.].
- [37] Канадската строителна асоциация и KPMG в Канада, 2021 г., "Строителството в един цифров свят", 1 май 2021 г. [Онлайн]. Достъпно на: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Дата на достъп: 5 март 2025 г.].
- [38] ZCS, "Декодиране на петата индустриална революция". [Онлайн]. Достъпно на: <https://www.pwc.in/decoding-the-fifth-industrial-revolution.html>. [Дата на достъп: 15 март 2025 г.].
- [39] M. K, Private Rights and Public Problems: The Global Economics of, Peterson Inst. for Intern. Economics, 2012 г.
- [40] F. N. a. Y. Z. Harvard Business School: Мануел Хофман, "Стойността на софтуера с отворен код", 24 януари 2024 г. [Онлайн]. Достъпно на: <https://www.hbs.edu/faculty/Pages/item.aspx?num=65230>. [Дата на достъп: 15 март 2025 г.].
- [41] Naval Centre for Cost Analysis Air Force Cost Analysis Agency, "Software Development Cost Estimating Handbook", 1 септември 2008 г. [Онлайн]. Достъпно на: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/SW%20Cost%20Est%20Manual%20Vol%201%20rev%2010.pdf>.
- [42] McKinsey, "Подобряване на производителността в строителството", [Онлайн]. Достъпно на: <https://www.mckinsey.com/capabilities/operations/our-insights/improving-construction-productivity>. [Дата на обръщение: 15 март 2025 г.].
- [43] A. G. a. C. Сиверсън, "Странният и ужасен път на производителността в строителния сектор на САЩ", 19 януари 2023 г. [Онлайн]. Достъпно на: <https://bfi.uchicago.edu/insight/research-summary/the-strange-and-awful-path-of-productivity-in-the-us-construction-sector/>. [Дата на

- достъп: 1 март 2025 г.].
- [44] McKinsey, "Постигането на производителност в строителството вече не е по избор", 9 август 2024 г. [Онлайн]. Достъпно на: <https://www.mckinsey.com/capabilities/operations/our-insights/delivering-on-construction-productivity-is-no-longer-optional>. [Дата на достъп: 5 март 2025 г.].
- [45] ING Group, "Изоставането на производителността в строителството води до увеличаване на разходите за строителство", 12 декември 2022 г. [Онлайн]. Достъпно на: <https://think.ing.com/articles/lagging-productivity-drives-up-building-costs-in-many-eu-countries/>. [Дата на обръщение: 15 март 2025 г.].
- [46] М. Берман, "Шокираща прогноза на главния изпълнителен директор на Microsoft: "Агентите ще заменят ВСЕКИ софтуер", 19 декември 2024 г. [Онлайн]. Достъпно на: <https://www.youtube.com/watch?v=uGOLYz2pgr8>. [Дата на достъп: 15 март 2025 г.].
- [47] Business Insider, "Главният изпълнителен директор на Anthropic твърди, че след 3 до 6 месеца изкуственият интелект ще пише 90% от кода, за който са отговаряли разработчиците на софтуер", 15 март 2025 г. [Онлайн]. Достъпно на: <https://www.businessinsider.com/anthropic-ceo-ai-90-percent-code-3-to-6-months-2025-3>. [Дата на достъп: 30 март 2025 г.].
- [48] Statista, "Сравнение на популярността на системите за управление на бази данни (СУБД) в световен мащаб към юни 2024 г., по категории", юни 2024 г. [Онлайн]. Достъпно на: <https://www.statista.com/statistics/1131595/worldwide-popularity-database-management-systems-category/>. [Дата на адреса: 15 март 2025 г.].
- [49] DB-Engines, "Класация на DB-Engines", [Онлайн]. Достъпно на: <https://db-engines.com/en/ranking>. [Дата на справка: 15 март 2025 г.].
- [50] "Проучване на разработчиците на Stack Overflow за 2023 г.", 2024 г. [Онлайн]. Достъпно на: <https://survey.stackoverflow.co/2023/>.
- [51] "SQL", 2024 Г. [Онлайн]. Достъпно на: <https://en.wikipedia.org/wiki/SQL>.
- [52] "Структурирани и неструктурни данни: каква е разликата?", 2024 г. [Онлайн]. Достъпно на: <https://www.ibm.com/blog/structured-vs-unstructured-data/>.
- [53] DataDrivenConstruction, "PDF СЪПОРАЗУМЕНИЕ НА ФОРМАТИТЕ НА ДАННИТЕ ЗА СТРОИТЕЛНИТЕ ПРОЕКТИ", 23 април 2024 г. [Онлайн]. Достъпно на: <https://datadrivenconstruction.io/wp-content/uploads/2024/10/COMPARISON-OF-DATA-FORMATS-FOR-CONSTRUCTION-PROJECTS-1.pdf>.
- [54] "Сайт за информационно моделиране на сгради", 2003 г. [Онлайн]. Достъпно на: <https://web.archive.org/web/20030711125527/http://usa.autodesk.com/adsk/servlet/item?itemID=2255342&siteID=123112>.

- [55] А. Бойко, "Лобистки войни и развитие на BIM. Част 5: BlackRock е господар на всички технологии. Как корпорациите контролират кода с отворен код", 2024 г. [Онлайн]. Достъпно на: <https://bigdataconstruction.com/autodesk-oracle-blackrock-open-source/>.
- [56] Д. Ушаков, "Директното моделиране - кой и защо се нуждае от него? Преглед на конкурентните технологии", 14 11 2011 г. [Онлайн]. Достъпно на: [https://isicad.net/articles.php?article\\_num=14805](https://isicad.net/articles.php?article_num=14805). [Дата на обръщение: 02 2025 г.].
- [57] С. Ийстман и А. Кетърс, "Ийстман, Чарлз; И Кетърс", септември 1974 г. [Онлайн]. Достъпно на: <https://files.eric.ed.gov/fulltext/ED113833.pdf>. [Дата на достъп: 15 март 2025 г.].
- [58] Д. Ушаков, "Директно моделиране - кой и защо се нуждае от него? Преглед на конкурентните технологии", 11 ноември 2011 г. [Онлайн]. Достъпно на: [https://isicad.net/articles.php?article\\_num=14805](https://isicad.net/articles.php?article_num=14805). [Дата на позоваване: 15 март 2025 г.].
- [59] Д. Вайсберг, "История на CAD", 12 декември 2022 г. [Онлайн]. Достъпно на: [https://www.shapr3d.com/blog/history-of-cad?utm\\_campaign=cadhistorynet](https://www.shapr3d.com/blog/history-of-cad?utm_campaign=cadhistorynet). [Дата на позоваване: 15 март 2025 г.].
- [60] ADSK, "Бяла книга за информационно моделиране на сгради", 2002 г. [Онлайн]. Достъпно на: [https://web.archive.org/web/20060512180953/http://images.adsk.com/apac\\_sapac\\_main/files/4525081\\_BIM\\_WP\\_Rev5.pdf#expand](https://web.archive.org/web/20060512180953/http://images.adsk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand). [Дата на позоваване: 15 март 2025 г.].
- [61] ADSK, "Бяла книга "Информационно моделиране на сгради в практиката", [Онлайн]. Достъпно на: [https://web.archive.org/web/20060512181000/http://images.adsk.com/apac\\_sapac\\_main/files/4525077\\_BIM\\_in\\_Practice.pdf](https://web.archive.org/web/20060512181000/http://images.adsk.com/apac_sapac_main/files/4525077_BIM_in_Practice.pdf). [Дата на достъп: 15 март 2025 г.].
- [62] А. Бойко, "Лобистките войни и развитието на BIM. Част 2: отворен BIM срещу затворен BIM. Европа срещу останалата част от света", 2024 г. [Онлайн]. Достъпно на: <https://bigdataconstruction.com/lobbyist-wars-and-the-development-of-bim-part-2-open-bim-vs-closed-bim-revit-vs-archicad-and-europe-vs-the-rest-of-the-world/>.
- [63] A. Boyko, "Lobbykriege um Daten im Bauwesen | Techno-Feudalismus und die Geschichte von BIMs", 2024. [Онлайн]. Достъпно на: [https://youtu.be/S-TNdUgfHxk?si=evM\\_v28KQbGOG0k&t=1360](https://youtu.be/S-TNdUgfHxk?si=evM_v28KQbGOG0k&t=1360).
- [64] ADSK, "Whitepaper BIM", 2002 г. [Онлайн]. Достъпно на: [https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac\\_sapac\\_main/files/4525081\\_BIM\\_WP\\_Rev5.pdf#expand](https://web.archive.org/web/20060512180953/http://images.autodesk.com/apac_sapac_main/files/4525081_BIM_WP_Rev5.pdf#expand). [Дата на обръщение: 15 март 2025 г.].
- [65] ADSK, "Integrated Design-Through-Manufacturing: Benefits and Rationale" ("Интегрирано проектиране чрез производство: ползи и обосновка"), [Онлайн]. Достъпно на: [https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489\\_Benefits\\_of\\_MAI.pdf](https://web.archive.org/web/20010615093351/http://www3.adsk.com:80/adsk/files/734489_Benefits_of_MAI.pdf). [Дата на достъп: 15 март 2025 г.].

- [66] M. Shacklett, "Структурирани и неструктурни данни: 2024. [Онлайн]. Достъпно на: <https://www.datamation.com/big-data/structured-vs-unstructured-data/>.
- [67] К. Уулард, "Как да се справим с нарастването на неструктурните данни", 2024 г. [Онлайн]. Достъпно на: <https://automationhero.ai/blog/making-sense-of-the-rise-of-unstructured-data/>.
- [68] А. С. О. J. L. D. Дж. а. Л. Т. Г. Майкъл П. Галахър, "Анализ на разходите при недостатъчна оперативна съвместимост", 2004 г. [Онлайн]. Достъпно на: <https://nvlpubs.nist.gov/nistpubs/gcr/2004/nist.gcr.04-867.pdf>. [Дата на достъп: 02 2025 г.].
- [69] CrowdFlower, "Доклад за науката за данните за 2016 г.", 2016 г. [Онлайн]. Достъпно на: [https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower\\_DataScienceReport\\_2016.pdf](https://visit.figure-eight.com/rs/416-ZBE-142/images/CrowdFlower_DataScienceReport_2016.pdf). [Дата на достъп: 15 март 2025 г.].
- [70] Analyticsindiamag, "6 задачи, които отнемат най-много време на учените, занимаващи се с данни", 15 май 2019 г. [Онлайн]. Достъпно на: <https://analyticsindiamag.com/ai-trends/6-tasks-data-scientists-spend-the-most-time-doing/>.
- [71] BizReport, "Доклад: учените, занимаващи се с данни, прекарват голяма част от времето си в почистване", 06 юли 2015 г. [Онлайн]. Достъпно на: <https://web.archive.org/web/20200824174530/http://www.bizreport.com/2015/07/report-data-scientists-spend-bulk-of-time-cleaning-up.html>. [Дата на достъп: 5 март 2025 г.].
- [72] S. Хокинг, "Science AMA Series: Stephen Hawking AMA Answers!", 27 юли 2015 г. [Онлайн]. Достъпно на: [https://www.reddit.com/r/science/comments/3nyn5i/science\\_ama\\_series\\_stephen\\_hawking\\_ama\\_answers/](https://www.reddit.com/r/science/comments/3nyn5i/science_ama_series_stephen_hawking_ama_answers/). [Дата на достъп: 15 март 2025 г.].
- [73] Б. Сайфърс и К. Доктороу, "Поверителност без монопол: защита на данните и оперативна съвместимост", 2024 г. [Онлайн]. Достъпно на: <https://www.eff.org/wp/interoperability-and-privacy>.
- [74] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information" ("Отворени данни: Отключване на иновациите и ефективността с помощта на ликвидна информация"), 1 октомври 2013 г. [Онлайн]. Достъпно на: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>. [Дата на достъп: 15 март 2025 г.].
- [75] А. Бойко, "Борбата за отворени данни в строителния бранш. Историята на AUTOLISP, intelliCAD, openDWG, ODA и openCASCADE", 15 05 2024 г. [Онлайн]. Достъпно на: <https://boikoartem.medium.com/the-struggle-for-open-data-in-the-construction-industry-2b97200e6393>. [Дата на обръщение: 16 02 2025 г.].
- [76] Уикипедия, "Майкрософт и отвореният код", [Онлайн]. Достъпно на: [https://en.wikipedia.org/wiki/Microsoft\\_and\\_open\\_source](https://en.wikipedia.org/wiki/Microsoft_and_open_source). [Дата на достъп: 15 март 2025 г.].

- [77] ВРЕМЕ, "Разликата между отворените и затворените модели на изкуствен интелект може да се свие. Ето защо това е важно", 5 ноември 2024 г. [Онлайн]. Достъпно на: <https://time.com/7171962/open-closed-ai-models-epoch/>. [Дата на достъп: 15 март 2025 г.].
- [78] The Verge, "Повече от една четвърт от новия код в Google се генерира от изкуствен интелект", 29 октомври 2024 г. [Онлайн]. Available: <https://www.theverge.com/2024/10/29/24282757/google-new-code-generated-ai-q3-2024>. [Дата на достъп: 15 март 2025 г.].
- [79] McKinsey Digital, "Бизнес аргументи за използване на графични процесори за ускоряване на обработката на анализи", 15 декември 2020 г. [Онлайн]. Достъпно на: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/the-business-case-for-using-gpus-to-accelerate-analytics-processing>. [Достъп: 15 март 2025 г.].
- [80] PwC, "PwC Open Source Monitor 2019", 2019 г. [Онлайн]. Достъпно на: <https://www.pwc.de/de/digitale-transformation/open-source-monitor-research-report-2019.pdf>. [Достъпен на 15 март 2025 г.].
- [81] Тревърс Смит, "Откритата тайна: софтуер с отворен код", 2024 г. [Онлайн]. Достъпно на: <https://www.traverssmith.com/knowledge/knowledge-container/the-open-secret-open-source-software/>. [Дата на достъп: 15 март 2025 г.].
- [82] Делойт, "Процесът на прехвърляне на данни при корпоративни трансформации", 2021 г. [Онлайн]. Достъпно на: <https://www2.deloitte.com/content/dam/Deloitte/us/Documents/finance/us-the-data-transfer-process-in-corporate-transformations.pdf>. [Дата на позоваване: 15 март 2025 г.].
- [83] gov.uk, "Анализ на данни и изкуствен интелект при изпълнението на правителствени проекти", 20 март 2024 г. [Онлайн]. Достъпно на: <https://www.gov.uk/government/publications/data-analytics-and-ai-in-government-project-delivery/data-analytics-and-ai-in-government-project-delivery>. [Дата на достъп: 5 март 2025 г.].
- [84] "Цитат за произхода: Всичко трябва да бъде направено възможно най-просто, но не и по-просто", 13 май 2011 г. [Онлайн]. Достъпно на: <https://quoteinvestigator.com/2011/05/13/einstein-simple/>. [Дата на достъп: 15 март 2025 г.].
- [85] "Трансформатор (архитектура за дълбоко обучение)," [Онлайн]. Достъпно на: [https://en.wikipedia.org/wiki/Transformer\\_\(deep\\_learning\\_architecture\)](https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)). [Дата на достъп: 15 март 2025 г.].
- [86] "Статистика на изтеглянето на пакети Python", 2024 г. [Онлайн]. Достъпно на: <https://www.pepy.tech/projects/pandas>.
- [87] Interview Bit, "Топ 10 библиотеки на Python", 2023 г. [Онлайн]. Достъпно на: <https://www.interviewbit.com/blog/python-libraries/#:~:text=With%20more%20than%20137%2C000%20libraries,data%20manipulation%2C%20>

- and%20many%20more. [Дата на достъп: 30 март 2025 г.].
- [88] "NVIDIA и HP повишават ефективността на науката за данните и генеративния изкуствен интелект на работните станции", 7 март 2025 г. [Онлайн]. Достъпно на: <https://nvidianews.nvidia.com/news/nvidia-hp-supercharge-data-science-generative-ai-workstations>. [Дата на достъп: 15 март 2025 г.].
- [89] Р. Орак, "Как да обработим DataFrame с милиони редове за секунди", 2024 г. [Онлайн]. Достъпно на: <https://towardsdatascience.com/how-to-process-a-dataframe-with-millions-of-rows-in-seconds>.
- [90] Ç. Услу, "Какво е Kaggle?", 2024 г. [Онлайн]. Достъпно на: <https://www.datacamp.com/blog/what-is-kaggle>.
- [91] "Главният изпълнителен директор на NVIDIA Дженсън Хуанг - ключова лекция на COMPUTEX 2024", 2 юни 2024 г. [Онлайн]. Достъпно на: <https://www.youtube.com/live/pKXDVsWZmUU?si=Z3Rj1Las8wiPII2w>. [Достъпен на 15 март 2025 г.].
- [92] "Членове: учредители и корпоративни членове", 2024 г. [Онлайн]. Достъпно на: <https://www.opendesign.com/member-showcase>.
- [93] А. Бойко, "Ерата на промяната: IFC е в миналото или защо ADSK и други CAD доставчици са готови да се откажат от IFC за USD в 14 ключови факта", 24 ноември 2024 г. [Онлайн]. Достъпно на: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [Дата на достъп: 23 февруари 2025 г.].
- [94] А. Бойко, "Светът след BIM. Преход към данни и процеси и дали строителната индустрия се нуждае от семантика, формати и оперативна съвместимост", 20 дек. 2024 г. [Онлайн]. Достъпно на: <https://boikoartem.medium.com/the-post-bim-world-7e35b7271119>. [Дата на обръщение: 23 февруари 2025 г.].
- [95] N. I. o. Health, "NIH STRATEGIC PLAN FOR DATA SCIENCE", 2016 г. [Онлайн]. Достъпно на: [https://datascience.nih.gov/sites/default/files/NIH\\_Strategic\\_Plan\\_for\\_Data\\_Science\\_Final\\_508.pdf](https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf). [Дата на достъп: 23 февруари 2025 г.].
- [96] Harvard Business Review, "Bad Data Costs the U.S. \$3 Trillion Per Year" (Лошите данни струват на САЩ 3 трилиона долара годишно), 22 септември 2016 г. [Онлайн]. Достъпно на: <https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year>.
- [97] Delpha, "Въздействие на качеството на данните", 1 януари 2025 г. [Онлайн]. Достъпно на: <https://delpha.io/impacts-of-data-quality/>.
- [98] W. B. D. Ръководство, "Проектиране за поддържане: значението на съображенията за експлоатация и поддръжка по време на фазата на проектиране на строителни проекти". [Онлайн]. Достъпно на: <https://www.wbdg.org/resources/design-for-maintainability>. [Дата на

- достъп: 15 март 2025 г.].
- [99] O. o. D. C. P. a. Oversight, "Ръководство за планиране на превенция и контрол на корозията за военни системи и оборудване", април 2014 г. [Онлайн]. Достъпно на: <https://www.dau.edu/sites/default/files/Migrated/CopDocuments/CPC%20Planning%20Guidebook%204%20Feb%2014.pdf>. [Дата на достъп: 15 март 2025 г.].
- [100] Gartner, "Качество на данните: най-добри практики за точни прозрения", 1 януари 2025 г. [Онлайн]. Достъпно на: <https://www.gartner.com/en/data-analytics/topics/data-quality>.
- [101] "В търсene на пирон" [Онлайн]. Достъпно на: [https://en.wikipedia.org/wiki/For\\_Want\\_of\\_a\\_Nail](https://en.wikipedia.org/wiki/For_Want_of_a_Nail). [Дата на достъп: 15 март 2025 г.].
- [102] McKinsey Global Institute, "Open data: Unlocking innovation and performance with liquid information" ("Отворени данни: Отключване на иновациите и ефективността с помощта на ликвидна информация"), октомври 2013 г. [Онлайн]. Достъпно на: [https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi\\_open\\_data\\_fullreport\\_oct2013.pdf](https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf). [Дата на достъп: 15 март 2025 г.].
- [103] EY, "Пътят към въглеродна неутралност", 10 март 2023 г. [Онлайн]. Достъпно на: [https://www.ey.com/ru\\_kz/services/consulting/the-path-to-carbon-neutrality](https://www.ey.com/ru_kz/services/consulting/the-path-to-carbon-neutrality). [Дата на достъп: 15 март 2025 г.].
- [104] PWC, "ESG Awareness", 1 юли 2024 г. [Онлайн]. Достъпно на: <https://www.pwc.com/kz/en/assets/esg-awareness/kz-esg-awareness-rus.pdf>. [Дата на достъп: 15 март 2025 г.].
- [105] G. Хамънд, "Въплътеният въглерод - инвентаризация на въглерода и енергията (ICE)", 2024 г. [Онлайн]. Достъпно на: <https://greenbuildingencyclopaedia.uk/wp-content/uploads/2014/07/Full-BSRIA-ICE-guide.pdf>.
- [106] "CO2\_изчисляване на въплътения въглерод", 2024 г. [Онлайн]. Достъпно на: [https://github.com/datadrivenconstruction/CO2\\_calculating-the-embodied-carbon](https://github.com/datadrivenconstruction/CO2_calculating-the-embodied-carbon).
- [107] McKinsey, "Представяне на цифровото бъдеще на строителството", 24 юни 2016 г. [Онлайн]. Достъпно на: <https://www.mckinsey.com/capabilities/operations/our-insights/imagining-constructions-digital-future>. [Дата на достъп: 25 февруари 2025 г.].
- [108] Bund der Steuerzahler Deutschland e.V., "Das Schwarzbuch", 10 октомври 2024 г. [Онлайн]. Достъпно на: <https://steuerzahler.de/aktuelles/detail/das-schwarzbuch-202425/>. [Дата на обръщение: 15 март 2025 г.].
- [109] SAS, "Data lake and data warehouse - know the difference" ("Езеро от данни и склад за данни - познайте разликата"), [Онлайн]. Достъпно на: [https://www.sas.com/en\\_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html](https://www.sas.com/en_is/insights/articles/data-management/data-lake-and-data-warehouse-know-the-difference.html). [Дата на справка: 15 март

2025 г.].

- [110] ADSK, "Информационно моделиране на сгради", 2002 г. [Онлайн]. Достъпно на: [https://www.laiserin.com/features/bim/autodesk\\_bim.pdf](https://www.laiserin.com/features/bim/autodesk_bim.pdf). [Дата на обръщение: 15 март 2025 г.].
- [111] А. Бойко, "Историческа карта на BIM", 2024 г. [Онлайн]. Достъпно на: <https://bigdataconstruction.com/history-of-bim/>.
- [112] А. С. Борковски, "Определения на BIM по организации и стандарти", 27 декември 2023 г. [Онлайн]. Достъпно на: <https://encyclopedia.pub/entry/53149>. [Дата на обръщение: 5 март 2025 г.].
- [113] CAD Vendor, "OPEN BIM Programme", 2012 г. [Онлайн]. Достъпно на: [https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open\\_bim/](https://web.archive.org/web/20140611075601/http://www.graphisoft.com/archicad/open_bim/). [Дата на справка: 30 март 2025 г.].
- [114] Уикипедия, "Industry Foundation Classes," [Онлайн]. Достъпно на: [https://en.wikipedia.org/wiki/Industry\\_Foundation\\_Classes](https://en.wikipedia.org/wiki/Industry_Foundation_Classes). [Дата на достъп: 15 март 2025 г.].
- [115] Уикипедия, "IGES," [Онлайн]. Достъпно на: <https://en.wikipedia.org/wiki/IGES>. [Дата на позоваване: 30 март 2025 г.].
- [116] А. Бойко, "История на CAD (BIM)", 15 декември 2021 г. [Онлайн]. Достъпно на: [https://miro.com/app/board/o9J\\_lamL2cs=/](https://miro.com/app/board/o9J_lamL2cs=/). [Дата на обръщение: 24 февруари 2025 г.].
- [117] T. K. K. A. O. F. B. C. E. L. H. H. H. E. L. P. N. S. H. T. J. v. L. H. H. G. D. H. T. K. C. L. A. W. J. S. Francesca Noardo, "Reference study of IFC software support: the GeoBIM benchmark 2019 - Part I", 8 януари 2021 г. [Онлайн]. Достъпно на: <https://arxiv.org/pdf/2007.10951.pdf>. [Дата на достъп: 5 март 2025 г.].
- [118] И. Рогачев, "Да поговорим за BIM: Максим Нечипоренко | Renga | IFC | Domestic BIM", 13 април 2021 г. [Онлайн]. Достъпно на: <https://www.youtube.com/watch?t=3000&v=VO3Y9uzF9M&feature=youtu.be>. [Дата на обръщение: 5 март 2025 г.].
- [119] D. Apec, "RETS in Real Estate: Why It's Crucial for Efficiency & Growth" ("RETS в областта на недвижимите имоти: защо е от решаващо значение за ефективността и растежа"), 17 декември 2024 г. [Онлайн]. Достъпно на: <https://www.realpha.com/blog/rets-importance-in-real-estate-explained>. [Дата на достъп: 5 март 2025 г.].
- [120] "Flex token cost", 2024 г. [Онлайн]. Достъпно на: <https://www.adsk.com/buying/flex?term=1-YEAR&tab=flex>.
- [121] А. Бойко, "Forget BIM and democratise access to data (17. Kolloquium Investor - Hochschule - Bauindustrie)", 2024 г. [Онлайн]. Достъпно на: <https://www.bim.bayern.de/wp-content/uploads/2023/06/Kolloquium-17-TUM-Baprozessmanagement-und-Bay-Bauindustrie.pdf>.

- [122] Д. Хил, Д. Фолдези, С. Ферер, М. Фридман, Е. Лох и Ф. Плашке, "Решаване на загадката за производителността в строителството", 2015 г. [Онлайн]. Достъпно на: <https://www.bcg.com/publications/2015/engineered-products-project-business-solving-construction-industries-productivity-puzzle>.
- [123] "SCOPE - Projektdatenumgebung und Modellierung multifunktionaler Bauprodukte mit Fokus auf die Gebäudehülle", 1 януари 2018 г. [Онлайн]. Достъпно на: <https://www.ise.fraunhofer.de/de/forschungsprojekte/scope.html>. [Дата на достъп: 2 март 2025 г.].
- [124] Apple.com, "Pixar, Adobe, Apple и NVIDIA създават Алианс за OpenUSD, за да наложат отворени стандарти за 3D съдържание", 1 август 2023 г. [Онлайн]. Достъпно на: <https://www.apple.com/newsroom/2023/08/pixar-adobe-apple-adsk-and-nvidia-form-alliance-for-openusd/>. [Достъп: 2 март 2025 г.].
- [125] АЕСмаг, "Стратегия за гранулирани данни на ADSK", 25 юли 2024 г. [Онлайн]. Достъпно на: <https://aescmag.com/technology/autodesks-granular-data-strategy/>. [Дата на достъп: 15 март 2025 г.].
- [126] А. Бойко, "Ерата на промяната: IFC е нещо от миналото или защо ADSK и други CAD доставчици са готови да се откажат от IFC за USD в 14 ключови факта", 24 11 2024. [Онлайн]. Достъпно на: <https://boikoartem.medium.com/the-age-of-change-ifc-is-a-thing-of-the-past-or-why-adsk-and-other-cad-vendors-are-willing-to-3f9a82cccd10a>. [Дата на обръщението: 23 февруари 2025 г.].
- [127] А. Бойко, "ENG BIM Cluster 2024 | Битката за данни и приложението на LLM и ChatGPT в строителството", 7 август 2024 г. [Онлайн]. Достъпно на: ENG BIM Cluster 2024 | The Battle for Data and Application of LLM and ChatGPT in the Construction ("Битката за данните и приложението на LLM и ChatGPT в строителството"). [Достъпно на: 15 март 2025 г.].
- [128] "Джефри Зелдман представя", 6 май 2008 г. [Онлайн]. Достъпно на: <https://zeldman.com/2008/05/06/content-precedes-design/>. [Дата на достъп: 15 март 2025 г.].
- [129] А. Бойко, "Анализ на DWG с ChatGPT | DataDrivenConstruction", 5 март 2024 г. [Онлайн]. Достъпно на: <https://www.kaggle.com/code/artemboiko/dwg-analyse-with-chatgpt-datadrivenconstruction>. [Дата на справка: 15 март 2025 г.].
- [130] McKinsey , "Ръководство на McKinsey за постигане на по-добра конкурентоспособност в ерата на цифровите технологии и изкуствения интелект", 2023 г. [Онлайн]. Достъпно на: <https://www.mckinsey.com/featured-insights/mckinsey-on-books/rewired>. [Дата на обръщение: 30 март 2025 г.].
- [131] Forbes, "Data Storytelling: The Essential Data Science Skill Everyone Needs", 31 март 2016 г. [Онлайн]. Достъпно на: <https://www.forbes.com/sites/brentdykes/2016/03/31/data-storytelling-the-essential-data-science-skill-everyone-needs/>. [Дата на достъп: 15 март 2025 г.].

- [132] Ј. Бертин, "Графика и графична обработка на информацията", 8 септември 2011 г. [Онлайн]. Достъпно на: [https://books.google.de/books/about/Graphics\\_and\\_Graphic\\_Information\\_Process.html?id=csqX\\_xnm4tcC&redir\\_esc=y](https://books.google.de/books/about/Graphics_and_Graphic_Information_Process.html?id=csqX_xnm4tcC&redir_esc=y). [Дата на достъп: 15 март 2025 г.].
- [133] CauseWeb, "Wells/Wilks on Statistical Thinking", [Онлайн]. Достъпно на: <https://www.causeweb.org/cause/resources/library/r1266>. [Дата на достъп: 15 март 2025 г.].
- [134] Ministrymagazine, "Как науката откри Сътворението", януари 1986 г. [Онлайн]. Достъпно на: <https://www.ministrymagazine.org/archive/1986/01/how-science-discovered-creation>. [Дата на справка: 15 март 2025 г.].
- [135] BCG, "Data-Driven Transformation: Accelerate at Scale Now", 23 май 2017 г. [Онлайн]. Достъпно на: <https://www.bcg.com/publications/2017/digital-transformation-transformation-data-driven-transformation>. [Достъпно на 15 май 2024 г.].
- [136] "Как да изградим архитектура на данните, която да стимулира иновациите - днес и утре", 3 юни 2020 г. [Онлайн]. Достъпно на: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/how-to-build-a-data-architecture-to-drive-innovation-today-and-tomorrow>. [Дата на обръщение: 15 март 2025 г.].
- [137] Оксфорд, "Удро Уилсън 1856-1924". [Онлайн]. Available: <https://www.oxfordreference.com/display/10.1093/acref/9780191866692.001.0001/q-oro-ed6-00011630>. [Дата на обръщението: 15 март 2025 г.].
- [138] "Конвертори", 2024 г. [Онлайн]. Достъпно на: <https://datadrivenconstruction.io/index.php/convertors/>.
- [139] PWC, "Sizing the prize What's the real value of AI for your business and how can you capitalise?", 1 януари 2017 г. [Онлайн]. Достъпно на: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. [Дата на достъп: 18 февруари 2025 г.].
- [140] "Тръбопровод в строеж", 2024 г. [Онлайн]. Достъпно на: <https://datadrivenconstruction.io/index.php/pipeline-in-construction/>.
- [141] Уикипедия, "Apache NiFi", 1 януари 2025 г. [Онлайн]. Достъпно на: [https://de.wikipedia.org/wiki/Apache\\_NiFi](https://de.wikipedia.org/wiki/Apache_NiFi). [Дата на достъп: 5 март 2025 г.].
- [142] n8n, "Gmail AI Auto-Responder: Създаване на чернови на отговори на входящи имейли", 1 май 2024 г. [Онлайн]. Достъпно на: <https://n8n.io/workflows/2271-gmail-ai-auto-responder-create-draft-replies-to-incoming-emails/>. [Достъпен на 15 март 2025 г.].
- [143] n8n, "Automation Real Estate Daily Deals with Zillow API, Google Sheets and Gmail", 1 март 2025 г. [Онлайн]. Достъпно на: <https://n8n.io/workflows/3030-real-estate-daily-deals-automation-with->

- zillow-api-google-sheets-and-gmail/. [Дата на достъп: 15 март 2025 г.].
- [144] В. Т. О'Нийл, "Процентът на неуспешните проекти за анализ, изкуствен интелект и големи данни = 85% - ура!", 1 януари 2025 г. [Онлайн]. Достъпно на:  
<https://designingforanalytics.com/resources/failure-rates-for-analytics-bi-iot-and-big-data-projects-85-yikes/>.
- [145] J. Neyman, On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Oxford University Press, 1934 г.
- [146] Т. J. S. a. J. S. Jesse Perla, "A Problem that Stumped Milton Friedman", Quantitative Economics with Julia, 1 януари 2025 г. [Онлайн]. Достъпно на:  
[https://julia.quantecon.org/dynamic\\_programming/wald\\_friedman.html](https://julia.quantecon.org/dynamic_programming/wald_friedman.html). [Дата на достъп: 1 май 2024 г.].
- [147] Т. Ландсал-Уелфеър, Прогнозиране на текущото настроение на нацията, Significance, 2012 г.
- [148] А. Бойко, "Сан Франциско. Строителен сектор 1980-2019 г.", 2024 г. [Онлайн]. Достъпно на:  
<https://www.kaggle.com/search?q=San+Francisco.+Building+sector+1980-2019>.
- [149] А. Бойко, "Kaggle: RVT IFC подава 5000 проекта", 2024 г. [Онлайн]. Достъпно на:  
<https://www.kaggle.com/datasets/artemboiko/rvtifc-projects>.
- [150] CFMA, "Подготовка за бъдещето с помощта на свързаното строителство". [Онлайн].  
Достъпно на: <https://cfma.org/articles/preparing-for-the-future-with-connected-construction>.  
[Достърен на 15 март 2025 г.].
- [151] Cisco, "Проучване на Cisco показва, че близо три четвърти от проектите за интернет на нещата се провалят", 22 май 2017 г. [Онлайн]. Достъпно на:  
<https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2017/m05/cisco-survey-reveals-close-to-three-fourths-of-iot-projects-are-failing.html>.
- [152] "Условия, необходими за запазване на растителни вкаменелости", 2024 г. [Онлайн]. Достъпно на: <https://ucmp.berkeley.edu/IB181/VPL/Pres/PresTitle.html>.
- [153] "Финк от BlackRock за облигациите, сливанията и придобиванията, рецесията в САЩ и изборите: пълно интервю", 2023 г. [Онлайн]. Достъпно на:  
<https://www.bloomberg.com/news/videos/2023-09-29/blackrock-s-fink-on-m-a-recession-election-full-intv-video>.
- [154] cio, "12 известни бедствия, свързани с изкуствения интелект", 02 октомври 2024 г. [Онлайн].  
Достъпно на: <https://www.cio.com/article/190888/5-famous-analytics-and-ai-disasters.html>.  
[Дата на достъп: 15 март 2025 г.].
- [155] G. Каспаров, Deep Thinking, PublicAffairs, 2017.

- [156] Уикипедия, "Kaggle", 1 януари 2025 г. [Онлайн]. Достъпно на: <https://en.wikipedia.org/wiki/Kaggle>. [Дата на достъп: 15 март 2025 г.].
- [157] Kaggle, "Титаник - машинно обучение от катастрофата", 1 януари 2025 г. [Онлайн]. Достъпно на: <https://www.kaggle.com/competitions/titanic/overview>. [Достъпен на 10 март 2025 г.].
- [158] Ш. Джохри, "Създаване на ChatGPT: от данни към диалог", 2024 г. [Онлайн]. Достъпно на: <https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/>.
- [159] П. Домингош, "Няколко полезни неща, които трябва да знаем за машинното обучение", 2024 г. [Онлайн]. Достъпно на: <https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf>.
- [160] J. Сарамаго, "Цитати за цитиране", [Онлайн]. Достъпно на: <https://www.goodreads.com/quotes/215253-chaos-is-merely-order-waiting-to-be-deciphered>. [Дата на достъп: 17 март 2025 г.].
- [161] NVIDIA, "Подобрете данните си за обучение с новите модели на класификатора NVIDIA NeMo Curator", 19 декември 2024 г. [Онлайн]. Достъпно на: <https://developer.nvidia.com/blog/enhance-your-training-data-with-new-nvidia-nemo-curator-classifier-models/>. [Достъпно на 25 март 2025 г.].
- [162] "NVIDIA обявява голямо издание на моделите на Фондация "Космос" и инструментите за физически данни за изкуствен интелект", 18 март 2025 г. [Онлайн]. Достъпно на: <https://nvidianews.nvidia.com/news/nvidia-announces-major-release-of-cosmos-world-foundation-models-and-physical-ai-data-tools>. [Дата на достъп: 25 март 2025 г.].
- [163] NVIDIA, "NVIDIA Isaac Sim", [в интернет]. Достъпно на: <https://developer.nvidia.com/isaac/sim>. [Достъпен на 25 март 2025 г.].
- [164] M. Quarterly, "Защо цифровите стратегии се провалят", 25 януари 2018 г. [Онлайн]. Достъпно на: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/why-digital-strategies-fail>. [Дата на достъп: 15 март 2025 г.].
- [165] М. J. Пери, "Любимите ми цитати на Милтън Фридман", 17 ноември 2006 г. [Онлайн]. Достъпно на: <https://www.aei.org/carpe-diem/my-favorite-milton-friedman-quotes/>. [Дата на достъп: 1 март 2025 г.].
- [166] J. A. Уилър, "Информация, физика, квант: търсене на връзки", 1990 г.
- [167] А. Бойко, "Лобистките войни и развитието на BIM. Част 5: BlackRock е господар на всички технологии. Как корпорациите контролират кода с отворен код", 2024 г. [Онлайн]. Достъпно на: <https://boikoartem.medium.com/lobbyist-wars-and-the-development-of-bim-d72ad0111a7d>.
- [168] T. Krijnen и J. Beetz, "A SPARQL query engine for binary-formatted IFC building models", *Advanced Engineering Informatics*, 2024.

- [171] "Брой предприятия в строителния сектор в Обединеното кралство през 2021 г., по размер на предприятието", 2024 г. [Онлайн]. Достъпно на: <https://www.statista.com/statistics/677151/uk-construction-businesses-by-size/>.
- [172] "5 000 проекта на МФК и РВТ", 2024 г. [Онлайн]. Достъпно на: <https://www.kaggle.com/code/artemboiko/5000-projects-ifc-rvt-datadrivenconstruction-io>.
- [173] М. Попова, "Това от бит: пионерът в областта на физиката Джон Арчибалд Уилър за информацията, природата на реалността и защо живеем във Вселена на участието", 2008 г. [Онлайн]. Достъпно на: <https://www.themarginalian.org/2016/09/02/it-from-bit-wheeler/>. [Дата на справка: февруари 2025 г.].
- [174] *Лобистки войни за данните в строителството / Технофеодализъм и историята на скритото минало на BIM*. [Филм]. Германия: Артьом Бойко, 2023 г.
- [175] А. Бойко, "CHATGPT WITH REVIT AND IFC | Автоматично извличане на документи и данни от проекти", 16 ноември 2023 г. [Онлайн]. Достъпно на: [https://www.youtube.com/watch?v=ASXolti\\_YPs&t](https://www.youtube.com/watch?v=ASXolti_YPs&t). [Дата на достъп: 2 март 2025 г.].
- [176] M. & Company, "Three new mandates for capturing a digital transformation's full value" ("Три нови мандата за получаване на пълната стойност на цифровата трансформация"), 22 януари 2022 г. [Онлайн]. Достъпно на: <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/three-new-mandates-for-capturing-a-digital-transformations-full-value>. [Дата на достъп: 15 февруари 2025 г.].
- [177] KPMG, "Строителството в дигиталния свят", 1 май 2021 г. [Онлайн]. Достъпно на: <https://assets.kpmg.com/content/dam/kpmg/ca/pdf/2021/05/construction-in-the-digital-age-report-en.pdf>. [Дата на достъп: 5 април 2024 г.].
- [178] LLP, KPMG, "Cue Construction 4.0: Make-or-Break Time", 17 март 2023 г. [Онлайн]. Достъпно на: <https://kpmg.com/ca/en/home/insights/2023/05/cue-construction-make-or-break-time.html>. [Дата на достъп: 15 февруари 2025 г.].
- [179] О. Бизнес, "Сатя Надела разкрива "Как агентите с изкуствен интелект ще разрушат моделите на SaaS", 10 януари 2025 г. [Онлайн]. Достъпно на: <https://www.outlookbusiness.com/artificial-intelligence/microsoft-ceo-satya-nadella-reveals-how-ai-agents-will-disrupt-saas-models>. [Дата на достъп: 15 март 2025 г.].
- [180] Forbes, "Почистване на големи данни: най-времеемката и най-неприятната задача в областта на науката за данните, сочи проучване", 23 март 2016 г. [Онлайн]. Достъпно на: <https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/>. [Дата на достъп: 15 март 2025 г.].
- [181] Министерство на външните работи, Британската общност и развитието на Обединеното кралство, "Стратегия за цифрово развитие от 2024 до 2030 г.", 18 март 2024 г. [Онлайн]. Достъпно на: <https://www.gov.uk/government/publications/digital-development-strategy-2024-to-2030>

- 2030/digital-development-strategy-2024-to-2030. [Дата на достъп: 15 март 2025 г.].
- [182] "Визия и стратегия в индустрията за проектиране на сгради", 7 ноември 2003 г. [Онлайн].  
Достъпно на:  
<https://web.archive.org/web/20030711125527/http://usa.adsk.com/adsk/servlet/item?id=2255342&siteID=123112>. [Дата на обръщение: 5 март 2025 г.].
- [183] М. Бочаров, "Информационно моделиране", март 2025 г. [Онлайн]. Достъпно на:  
<https://www.litres.ru/book/mihail-evgenovich-bocharov/informacionnoe-modelirovanie-v-rossii-71780080/chitat-onlayn/?page=5>. [Дата на обръщение: 15 март 2025 г.].
- [184] "Интегрирано проектиране чрез производство: ползи и обосновка", 2000 г. [Онлайн].  
Достъпно на:  
[https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489\\_Benefits\\_of\\_MAI.pdf](https://web.archive.org/web/20010615093351/http://www3.autodesk.com:80/adsk/files/734489_Benefits_of_MAI.pdf). [Дата на позоваване: 25 март 2025 г.].
- [185] CAD Vendor, "Програмата Open BIM е маркетингова кампания", 12 март 2012 г. [Онлайн].  
Достъпно на:  
<https://web.archive.org/web/20120827193840/http://www.graphisoft.com/openbim/>. [Дата на обръщение: 30 март 2025 г.].

# ИНДЕКС НА ПРЕДМЕТИТЕ

3D, 8, 14, 71, 73, 84, 191, 210, 215, 232, 234, 263, 276, 277, 287, 298, 299, 302, 303, 306, 307, 337, 375, 393, 394, 448, 468, 480

4D, 84, 172, 196, 199, 210, 229, 234, 237, 287  
4IR, 43

5D, 84, 172, 196, 210, 237, 287, 492

6D, 172, 196, 229, 232, 233, 234, 235

7D, 232, 233, 234, 287

8D, 172, 196, 229, 232, 233, 234, 287

## A

AI, 3, 50, 52, 100, 102, 103, 106, 107, 116, 395, 457, 459, 461, 477

AIA, 289

AIM, 3, 289

AMS, 14, 84, 151, 153

Apache Airflow, 311, 361, 362, 363, 364, 366, 367, 369, 371, 399, 494

Apache NiFi, 116, 311, 361, 367, 368, 371, 399, 482, 494

Апачи ORC, 62, 330, 378

Apache Parquet, 62, 67, 380, 381, 494

API, 54, 94, 95, 97, 109, 137, 138, 154, 168, 218, 219, 256, 257, 260, 271, 273, 294, 295, 296, 297, 300, 308, 326, 329, 342, 366, 369, 370, 488

## В

БДС, 258, 259

Големи данни, 9, 67, 245, 383

BIM, 2, 3, 4, 6, 3, 14, 17, 24, 56, 58, 60, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 84, 92, 137, 138, 139, 140, 141, 142, 144, 150, 154, 156, 166, 172, 183, 184, 186, 187, 190, 195, 196, 208, 210, 211, 213, 216, 217, 218, 219, 220, 221, 222, 227, 228, 237, 238, 239, 242, 243, 246, 250, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 266, 271, 272, 273, 275, 276, 277, 278, 279, 280, 282, 285, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 308, 309, 324, 328, 337, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 406, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

475, 476, 492, 494, 497

BlackBox, 240, 242, 243.

BMS, 8

Боке, 320, 337

BOM, 76, 77, 79, 257, 263

Ограничителна кутия, 234, 373, 392, 393, 394, 414, 415, 480

BREP, 142, 234, 263, 264, 276, 283, 284

## C

CAD, 6, 14, 18, 24, 56, 57, 58, 63, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 84, 85, 95, 95, 97, 111, 126, 137, 138, 139, 140, 141, 142, 144, 146, 147, 152, 153, 155, 156, 166, 172, 175, 183, 184, 186, 187, 190, 195, 196, 206, 208, 210, 211, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 224, 227, 228, 232, 234, 237, 238, 239, 242, 243, 251, 252, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 303, 308, 309, 324, 328, 337, 344, 351, 356, 358, 361, 367, 372, 376, 383, 389, 402, 403, 405, 406, 413, 416, 431, 448, 462, 466, 475, 476, 480, 492, 494, 497

CAE, 16, 283

CAFM, 14, 24, 62, 84, 151, 153, 172, 233, 278, 326, 356, 387, 458, 473

CAM, 16, 78

CAPEX, 14, 82

CDE, 84, 175, 388, 389, 390

ChatGPT, 103, 104, 109, 110, 124, 129, 221, 303, 307, 343, 349, 421

Клод, 103, 104, 109, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

CO<sub>2</sub>, 72

CO<sub>2</sub>, 229, 234, 235, 236, 237, 238, 239, 243

COBie, 156, 289, 292.

CE, 56, 168, 169, 170, 477

Копилот, 114, 116, 459

CPIXML, 143, 272, 273, 276, 277, 279, 285, 296

CPM, 14, 17, 62, 166, 175, 233, 326, 473

CQMS, 14, 84, 177, 178, 462

CRM, 109, 369, 459.

CRUD, 51, 459

CSG, 263

CSV, 61, 62, 88, 89, 120, 123, 128, 129, 130, 131, 135, 168, 268, 272, 280, 329, 333, 344, 354, 354, 356, 373, 377, 378, 379, 380, 401, 407, 414, 474, 480

**D**

DAE, 276, 277, 278, 280, 281, 284, 285, 414  
DAG, 362, 363, 365, 366  
Dash, 320, 336, 337.  
Управление на данните, 373, 395, 396, 398, 400, 401  
Езеро от данни, 214, 373, 376, 384, 385, 386, 387, 388, 389, 390, 400  
Data Lakehouse, 67, 373, 386, 387, 388  
Минимализъм на данните, 373, 395, 396, 400, 401  
Data Swamp, 373, 395, 397, 401  
Хранилище за данни, 382, 383, 400  
Данни като услуга, 487  
базирани на данни, 50, 170, 460, 461, 484, 486  
DataFrame, 67, 117, 121, 122, 123, 125, 129, 130, 131, 133, 134, 135, 137, 219, 220, 224, 237, 308, 328, 329, 330, 332, 333, 344, 345, 347, 348, 354, 365, 377, 381, 409, 414, 415, 475  
DataOps, 170, 373, 398, 399, 400, 401, 475  
DeepSeek, 103, 104, 107, 109, 110, 124, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494  
DGN, 8, 140, 186, 227, 357  
DWG, 8, 70, 71, 73, 97, 140, 186, 211, 227, 272, 287, 302, 303, 304, 307, 357, 376, 497  
DWH, 67, 373, 376, 382, 383, 384, 386, 387, 388, 389, 390  
DXF, 8, 73, 277

**E**

ECM, 58, 175  
ECS, 142  
EIR, 289  
eLOD, 289  
ELT, 384, 385  
EPM, 14, 166, 198  
ERP, 2, 11, 12, 14, 17, 18, 24, 25, 58, 62, 109, 153, 166, 172, 175, 196, 198, 210, 232, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 272, 277, 278, 279, 282, 326, 351, 356, 361, 369, 387, 388, 389, 390, 458, 462, 473, 475, 487, 492  
ESG, 196, 235, 236, 238  
ETL, 6, 1.1-8, 32, 49, 81, 113, 116, 119, 128, 188, 193, 219, 291, 311, 312, 317, 323, 324, 325, 326, 327, 328, 330, 331, 333, 338, 339, 340, 343, 344, 348, 349, 350, 351, 353, 354, 356, 361, 362, 363, 364, 365, 367, 371, 372, 381, 382, 383, 384, 385, 399, 409, 430, 475, 481, 494  
Excel, 57, 61, 62, 65, 66, 85, 88, 111, 120, 123, 123, 125, 132, 154, 167, 187, 210, 223, 224, 226, 227, 228, 278, 291, 329, 333, 342, 351, 365, 376, 378, 459, 475, 497  
Извлечение, 81, 128, 134, 193, 311, 323, 324, 326, 328, 330, 345, 361, 362, 363, 365, 383, 384, 450, 475

**F**

Перо, 62, 123, 330, 378  
FPDF, 339, 340, 341, 342, 343  
GDPR, 109  
ГИС, 58  
GLTF, 143, 278  
Таблицы на Google, 368, 370  
Grok, 103, 104, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 429, 443

**H**

HDF5, 62, 67, 123, 329, 330, 378, 379, 380  
HiPPO, 29, 37, 95, 424, 477, 484, 490  
HTML, 123, 340, 365, 370

**I**

IDS, 289, 290, 291.  
IFC, 8, 73, 138, 142, 186, 227, 261, 262, 263, 264, 265, 266, 267, 268, 272, 273, 273, 276, 277, 278, 279, 280, 284, 286, 292, 296, 302, 329, 357, 414, 415, 417, 497  
IGES, 262, 263, 276.  
iLOD, 289  
IoT, 10, 18, 67, 271, 367, 369, 405, 413, 417, 417, 418, 419, 455, 460, 465, 482, 484  
ISO 19650, 388

**J**

JavaScript, 320, 369, 378  
JSON, 88, 89, 90, 92, 123, 128, 142, 269, 272, 280, 329, 330, 333, 378, 474, 480  
Jupyter Notebook, 114, 115, 116, 130, 187, 224, 330, 346, 417, 425

**K**

Kaggle, 115, 121, 130, 187, 224, 303, 307, 330, 346, 408, 415, 417, 425, 426, 430, 431, 433  
k-NN, 392, 393, 442, 445, 446, 447, 448  
KPI, 245, 311, 317, 318, 319, 320, 321, 324, 353, 372, 389, 478, 479

**L**

LEED, 235, 236, 238.  
LlaMa, 103, 104, 120, 124, 129, 133, 166, 219, 300, 329,

343, 349, 381, 421, 429, 443, 494

LLM, 3, 4, 24, 29, 50, 51, 52, 55, 56, 92, 95, 99, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 118, 120, 124, 125, 126, 129, 131, 133, 134, 135, 166, 166, 187, 219, 220, 220, 221, 222, 222, 224, 225, 226, 231, 237, 238, 251, 294, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 328, 329, 332, 333, 334, 335, 336, 338, 339, 340, 343, 344, 345, 346, 347, 348, 349, 354, 370, 372, 376, 381, 382, 392, 394, 399, 413, 425, 429, 430, 432, 433, 435, 443, 444, 457, 460, 461, 476, 488, 494

Натоварване, 81, 128, 134, 193, 311, 323, 325, 326, 333, 334, 338, 339, 342, 343, 361, 362, 363, 365, 367, 383, 384, 450, 475

LOD, 287, 289

LOI, 287

LOMD, 287

Low-Code, 368, 369

## M

Matplotlib, 123, 306, 320, 333, 335, 342, 372, 409, 411, 494

MCAD, 77, 78, 257, 284

MEP, 14, 175, 492

МРЕЖА, 234, 283, 284, 285, 296, 462

Microsoft SQL, 65

Mistral, 103, 104, 107, 110, 129, 133, 166, 219, 221, 300, 329, 343, 349, 381, 421, 429, 443, 494

MRP, 11, 12

MS Project, 70

MySQL, 63, 64, 65, 291, 329, 330

## N

n8n, 116, 311, 361, 368, 369, 370, 371, 372

NLP, 69, 394

No-Code, 368, 369

NURBS, 142, 282, 283, 284, 285

NWC, 8, 276

## O

OBJ, 143, 273, 276, 277, 278, 280, 281, 284, 285

OCCT, 273

OCR, 69, 128, 132, 134

OmniClass, 154, 155, 156.

отворен BIM, 142, 216, 219, 256, 261, 278, 291

Отворен код, 45, 55, 97, 98, 107, 108, 273, 275, 494

COBA, 267, 268, 269.

## P

Панди, 56, 67, 103, 117, 118, 119, 120, 121, 122, 123, 125, 130, 134, 138, 186, 187, 220, 221, 225, 227, 269, 300, 303, 328, 329, 330, 377, 378, 380, 381, 403, 408, 409, 410, 411, 412, 414, 475, 479, 494

Паркет, 67, 123, 329, 330, 373, 378, 379, 380, 381, 382, 386, 401, 403, 414, 474, 480

PDF, 69, 70, 71, 85, 111, 126, 127, 128, 129, 130, 131, 132, 134, 146, 147, 177, 186, 190, 191, 211, 215, 278, 338, 339, 340, 341, 342, 342, 344, 345, 346, 347, 354, 355, 356, 357, 365, 376, 475, 480, 497

PDM, 16

PHP, 63, 339

Тръбопровод, 44, 53, 115, 128, 183, 303, 307, 308, 311, 312, 349, 350, 351, 352, 354, 355, 356, 367, 370, 417, 430, 476

PLM, 16, 246

PLN, 8, 272, 296

Plotly, 320, 336, 337, 372

PMIS, 3, 24, 32, 62, 152, 196, 198, 211, 239, 240, 245, 246, 247, 248, 249, 250, 272, 326, 351, 387, 388, 389, 390, 458, 462, 487

PMS, 84, 151, 233

PostgreSQL, 63, 64, 65, 329, 395

Power BI, 320, 372, 482

частен капитал, 464, 487

Python, 56, 63, 103, 105, 112, 114, 115, 117, 118, 119, 129, 130, 131, 133, 134, 135, 166, 167, 179, 187, 219, 224, 225, 227, 303, 304, 308, 320, 329, 330, 332, 333, 339, 340, 346, 348, 356, 367, 369, 372, 378, 409, 417, 426, 459, 475, 476, 479, 494

## Q

QTO, 72, 196, 214, 215, 216, 217, 218, 219, 221, 223, 225, 226, 228, 237, 238, 242, 243, 251, 301, 302, 475, 479

QWEN, 103, 104, 124, 129, 133, 166, 219, 221, 300, 329, 349, 343, 349, 381, 429, 443

## R

RAG, 111, 116

RDBMS, 63, 64, 65, 82, 89.

RDF, 267, 268, 269

RegEx, 126, 136, 177, 179, 331, 332, 333, 357, 474

RFID, 8, 18, 58, 84, 418, 419, 455, 460, 465, 482

ROI, 311, 317, 319, 321, 351, 370, 389, 479

RPM, 14, 84, 151, 331, 333, 334

RVT, 8, 73, 77, 140, 186, 227, 272, 296, 300, 302, 322, 329, 357, 376, 414, 415, 417, 497

## S

SaaS, 24, 50, 51, 52, 458  
ОБХВАТ, 273, 277  
SDK, 139, 141, 257, 264, 273, 281, 286, 296, 329, 414  
Seaborn, 123, 320, 336, 337, 408, 412  
SPARQL, 269  
SQL, 63, 65, 66, 88, 89, 103, 105, 119, 123, 166, 168, 268, 269, 279, 276, 277, 300, 329, 376, 392, 494  
SQLite, 63, 64, 65, 166, 167, 296, 329, 363  
СТЪПКА, 261, 262, 263, 266, 272, 276, 277, 292  
Streamlit, 336, 337  
SVF, 142, 276, 285

## T

Трансформация, 128, 193, 311, 323, 325, 326, 330, 331, 333, 334, 340, 341, 342, 344, 347, 361, 362, 363, 365, 367, 383, 384, 450, 475

## U

Uniclass, 154, 155, 156.  
USD, 142, 143, 255, 276, 277, 278, 279, 280, 281, 284, 285, 286, 292, 296

## V

VectorOps, 373, 398, 400, 401  
VR, 84, 271, 285

## W

WhiteBox, 240, 242, 243.

## X

XLSX, 8, 61, 62, 123, 128, 129, 231, 268, 276, 277, 280, 296, 299, 308, 351, 373, 377, 378, 379, 380, 401, 474, 480  
XML, 61, 88, 89, 92, 128, 143, 269, 272, 277, 280, 280, 291, 299, 329, 378, 414, 474